

12-22-99

X

Please type a plus sign (+) inside this box → ☐

PTO/SB/05 (4/98)
Approved for use through 09/30/2000. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

12/21/99
Jc604 U.S. PTO

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 C.F.R. § 1.53(b))

Attorney Docket No.	6232.US.P1
First Inventor or Application Identifier	George G. Schlauder
Title	See 1 in Addendum
Express Mail Label No.	EL384204542US

OUT
P.S. N 9550
47397/60
C6/12/97

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

ADDRESS TO: Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

- ☒ * Fee Transmittal Form (e.g., PTO/SB/17)
(Submit an original and a duplicate for fee processing)
- ☒ Specification [Total Pages 121]
(preferred arrangement set forth below)
 - Descriptive title of the Invention
 - Cross References to Related Applications
 - Statement Regarding Fed sponsored R & D
 - Reference to Microfiche Appendix
 - Background of the Invention
 - Brief Summary of the Invention
 - Brief Description of the Drawings (if filed)
 - Detailed Description
 - Claim(s)
 - Abstract of the Disclosure
- ☒ Drawing(s) (35 U.S.C. 113) [Total Sheets 22]
- Oath or Declaration [Total Pages]
 - ☐ Newly executed (original or copy)
 - ☐ Copy from a prior application (37 C.F.R. § 1.63(d))
(for continuation/divisional with Box 16 completed)
 - ☐ DELETION OF INVENTOR(S)
Signed statement attached deleting inventor(s) named in the prior application, see 37 C.F.R. §§ 1.63(d)(2) and 1.33(b).

- ☐ Microfiche Computer Program (Appendix)
- Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)
 - ☒ Computer Readable Copy
 - ☒ Paper Copy (identical to computer copy)
 - ☒ Statement verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

- ☐ Assignment Papers (cover sheet & document(s))
- ☐ 37 C.F.R. §3.73(b) Statement of Power of Attorney (when there is an assignee)
- ☐ English Translation Document (if applicable)
- ☐ Information Disclosure Statement (IDS)/PTO-1449 [Copies of IDS Citations]
- ☐ Preliminary Amendment
- ☒ Return Receipt Postcard (MPEP 503)
(Should be specifically itemized)
- ☐ * Small Entity Statement(s) filed in prior application (PTO/SB/09-12) Status still proper and desired
- ☐ Certified Copy of Priority Document(s) (if foreign priority is claimed)
- ☒ Other: Application Cover Sheet

* NOTE FOR ITEMS 1 & 13 IN ORDER TO BE ENTITLED TO PAY SMALL ENTITY FEES, A SMALL ENTITY STATEMENT IS REQUIRED (37 C.F.R. § 1.27), EXCEPT IF ONE FILED IN A PRIOR APPLICATION IS RELIED UPON (37 C.F.R. § 1.28).

16. If a CONTINUING APPLICATION, check appropriate box, and supply the requisite information below and in a preliminary amendment:
☐ Continuation ☐ Divisional ☒ Continuation-in-part (CIP) of prior application No 09/173,141
Prior application information: Examiner Not assigned Group / Art Unit: Not assigned
For CONTINUATION or DIVISIONAL APPS only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 4b, is considered a part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.

17. CORRESPONDENCE ADDRESS

☐ Customer Number or Bar Code Label (Insert Customer No. or Attach bar code label here) or ☐ Correspondence address below

Name	Steven F. Weinstock Abbott Laboratories				
Address	Department 377 / AP6D-2 100 Abbott Park Road				
City	Abbott Park	State	IL	Zip Code	60064-6050
Country	USA	Telephone	(847) 938-3137	Fax	(847) 938-2623

Name (Print/Type)	Dianne Casuto	Registration No. (Attorney/Agent)	40,943
Signature	Dianne Casuto		Date December 21, 1999

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231

Addendum

1. METHODS AND COMPOSITIONS FOR DETECTING HEPATITIS E VIRUS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: George G. Schlauder, *et al.*

Serial No.:

Group No.:

Filed: December 21, 1999

Examiner:

For: METHODS AND COMPOSITIONS FOR DETECTING HEPATITIS E VIRUS

Box Patent Application

Assistant Commissioner for Patents

Washington, D.C. 20231

EXPRESS MAIL CERTIFICATE

"Express Mail" label number EL384204542US

Date of Deposit: 12/21/99

I hereby state that the following *attached* papers or fees

Utility Patent Application Transmittal— 2 pages

Fee Transmittal, 1 pg. (in duplicate)

Application Cover Sheet— 1 pg.

Specification (115 pages); Claims (5 pages); Abstract (1 page); **Total: 121**

Drawings (22 pages)

Paper copy of Sequence Listing— 140 pgs.

Computer Readable Form Copy of Sequence Listing— 1 disc

Statement to Support filings and Submission in Accordance with 37 CFR 1.821 through 1.825— 1 pg.

Two Return-Receipt Postcards

are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10, on the date indicated above and is addressed to the Box Patent Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

Wanda E. Smith



Signature of person mailing paper or fee

(Express Mail Certificate)

COVER SHEET

6232.US.P1

Filed: December 21, 1999

TITLE:

**METHODS AND COMPOSITIONS FOR DETECTING HEPATITIS E
VIRUS**

The inventors of the attached Patent Application are:

George G. Schlauder of Skokie, IL

James C. Erker of Grayslake, IL

Suresh M. Desai of Libertyville, IL

George J. Dawson of Libertyville, IL

Isa K. Mushahwar of Grayslake, IL 60030

Submitted by:

**Dianne Casuto, Counsel
Abbott Laboratories
Dept. 377, AP6D/2
100 Abbott Park Road
Abbott Park, IL. 60064-6050
(847) 938-3137**

METHODS AND COMPOSITIONS FOR DETECTING HEPATITIS E VIRUS

Related Applications

This application claims priority to U.S.S.N. 09/173,141, filed October 15, 1998, now pending, which claims priority under 35 U.S.C. §119(e) to provisional application U.S.S.N. 60/061,199, filed October 15, 1997, now abandoned, the disclosures of which are incorporated by reference herein.

Field of the Invention

This invention relates generally to methods and compositions for detecting hepatitis E virus, and more particularly to methods and compositions for detecting in, or treating individuals infected with US-type and US-subtype strains of hepatitis E virus.

Background of the Invention

There are at least five major classes of hepatotropic viruses that cause inflammation of the liver (hepatitis). These viruses include hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis D virus (HDV) and hepatitis E virus (HEV). Although only HBV, HCV and HDV cause chronic hepatitis, all five types cause acute disease either directly or as a result of superinfection/co-infection by, for example, HBV and HDV. HEV causes symptoms of hepatitis that are similar to those of other viral agents including abdominal pain, jaundice, malaise, anorexia, dark urine, fever, nausea and vomiting (see, for example, Reyes *et al.*, "Molecular biology of non-A, non-B hepatitis agents: hepatitis C and hepatitis E viruses" in *Advances in Virus Research* (1991) 40: 57-102; Bradley, "Hepatitis non-A, non-B viruses become identified as hepatitis C and E viruses" in *Progr. Med. Virol.* (1990) 37: 101-135; Hollinger "Non-A, non-B hepatitis viruses" in *Virology*, Second Edition (1990), Second Edition, Raven Press, New York pp. 2239-2271; Gust *et al.*, "Report of a workshop: waterborne non-A, non-B hepatitis" *J. Infect. Dis.* (1987) 156: 630-635; and Krawczynski "Hepatitis E" *Hepatology* (1993) 17: 932-941). Unlike the other hepatoviruses, however, HEV generally has not been perceived as being a significant cause of hepatitis in the US.

Geographic regions where HEV is endemic include eastern and northern Africa, India, Pakistan, Burma and China (Reyes *et al.* (1991) *supra*). The case fatality rate of HEV infection is estimated to be between about 0.1% to about 1.0% in the general population, where HEV is endemic, and as high as about 20% among pregnant women in developing countries. Most fatalities result from fulminant hepatitis (Reyes *et al.* (1991) *supra*). The occasional reports of infection with HEV in the US, western Europe and Japan, usually are observed in travelers returning home from visits to areas where HEV is endemic. However, there is little information pertaining to the morbidity and/or mortality of infection with HEV in the US since HEV infections are not reported to a central agency. Extensive, systematic studies have not been performed to determine the importance of HEV in US. Further, if such studies were performed, the relative importance of HEV in US (and possibly Japan and Western Europe) may continue to be underestimated unless the proper reagents are developed to conduct such a study.

The basic features of HEV is that it is a non-enveloped virus, approximately 27-30 nm in diameter possessing a positive sense, single stranded RNA genome which comprises three discontinuous open-reading frames (ORFs), referred to in the art as open reading frame 1 (ORF 1), open reading frame 2 (ORF 2), and open reading frame 3 (ORF 3). Based on the overall morphology of the virus and the size and organization of the genome, the virus is tentatively classified as a member of the Caliciviridae. The first two isolates of HEV to be identified and sequenced were obtained from Burma and from Mexico. The overall nucleic acid identity across the genome of both isolates is 76% (Reyes *et al.* (1990) *Science*, 247: 1335-1339; Tam *et al.* (1991) *Virology* 185: 120-131; Huang *et al.* (1992) *Virology* 191: 550-558). Many of the nucleotide differences were noted at the third codon position, such that the deduced similarities in amino acid sequences between the Burmese and Mexican strains of HEV were 83%, 93% and 87%, for open reading frames ORF 1, ORF 2, and ORF 3, respectively.

In the Burmese strain, there is a short non-translated region of about 27 nucleotides at the 5' end of the genome which has not been identified in the Mexican strain. ORF 1 comprises approximately 5,100 nucleotides, which encode several conserved motifs including a putative methyltransferase domain, an RNA helicase domain, a putative RNA-dependent RNA

polymerase (RDRP) domain, and a putative papain-like protease. A tripeptide sequence of Gly-Asp-Asp (GDD), found in all positive-sense RNA plant and animal viruses, is located within ORF 1 and usually signifies RDRP function. Conserved motifs suggestive of purine NTPases activity that is usually associated with cellular and viral helicases also are present in the ORF 1 sequence. There is no consistent immune response to gene products encoded in ORF 1.

The second open reading frame (ORF 2) occupies the carboxyl one-third of the viral genome. ORF 2 comprises approximately 2,000 nucleotides which encode a consensus signal peptide sequence at the amino terminus of ORF 2, and a putative capsid protein, translated in a 1+ reading frame in relation to ORF 1. Frequently, HEV infected individuals produce antibodies that react with peptides or recombinant proteins derived from ORF 2.

The third open reading frame (ORF 3) partly overlaps both ORF 1 and ORF 2, and comprises 369 nucleotides translated in the +2 reading frame in relation to ORF 1. Although the function of the protein encoded by ORF 3 is unknown, the protein is antigenic, with most HEV infected individuals producing antibodies to this protein. Accordingly, peptides or recombinant proteins derived from ORF 2 and ORF 3 may serve as serologic markers useful in diagnosing exposure to HEV.

Recently, several additional HEV isolates have been identified and compared to the Burmese and Mexican strains of HEV. Most of the recent isolates are more closely related to the Burmese strain than to the Mexican strain of HEV. Except for a brief appearance in 1986-1987, there have been no additional isolates of the Mexican strain of HEV (Velasquez *et al.* (1992) JAMA, 263: 3281-3286).

One isolate, referred to as SAR-55, recently was isolated from an HEV-infected individual from Pakistan. The SAR-55 isolate is highly related to the Burmese strain with nucleotide and amino acid identities of 94% and 99%, respectively, across the entire genome. Several other recent isolates have been made from the Chinese province of Xuar, bordering on Pakistan. These Chinese isolates were more closely related to the Pakistani strain (approximately 98% nucleotide identity) than to the Burmese strain (approximately 93% nucleotide identity).

Prior to the sequencing of the viral genome and the availability of viral-encoded recombinant proteins and synthetic peptides, HEV infection was monitored by electron microscopy and immunofluorescence. Soon after the identification of the HEV genome, specific laboratory techniques for detecting HEV infection became available including (i) specific immunoassays, for example, western blot assays and ELISA's based on recombinant proteins and/or synthetic peptides, and (ii) polymerase chain reactions (PCR), for example, reverse transcriptase PCR (RT-PCR). RT-PCR has been used successfully to detect HEV RNA in samples of stool or serum in cases of acute hepatitis infections, and in epidemics of ET-NANBH. Furthermore, by using recombinant antigens derived from the Mexican and Burmese strains of HEV, specific IgG, IgM and, in some cases, IgA antibodies to HEV have been detected in specimens obtained from ET-NANBH outbreaks in Somalia, Burma, Borneo, Tashkent, Kenya, Pakistan and Mexico. Specific IgG, and sometimes IgM antibodies to HEV have been detected in cases of acute, sporadic hepatitis in geographic regions such as Egypt, India, Tajikistan and Uzbekistan as well as in acute hepatitis cases among patients in industrialized nations (for example, US, UK, Netherlands and Japan) who traveled to areas endemic for HEV.

To date, PCR and immunoassay-based tests based on the Burmese and Mexican isolates of HEV have established that various cases of "waterborne hepatitis" were caused by HEV. The antibody tests also were important in establishing HEV as a cause of acute, sporadic hepatitis in developing nations and among travelers to regions where HEV is endemic. However, it is unclear as to how many cases of acute HEV currently go undiagnosed due to the inability of current reagents to detect exposure to all strains of HEV. Accordingly, as new isolates of HEV are identified, it is desirable to develop new compositions and methods for detecting and/or treating hepatitis caused by the new HEV strains, which heretofore remain undetectable by the currently available test kits.

Summary of the Invention

The invention is based, in part, upon the discovery of a new family of human hepatitis E viruses. The newly discovered family of hepatitis E viruses fall within a class referred to hereinafter as a US-type hepatitis E virus. Furthermore, two members of the family were

discovered in individuals living in the United States and exhibit considerable similarities when compared at the nucleotide and amino acid levels. The latter two members together belong to a subclass of the US-type hepatitis E virus, referred to hereinafter as US-subtype hepatitis E virus.

5 Accordingly, in one aspect, the invention provides a method for detecting the presence of a US-type or US-subtype hepatitis E virus in a test sample of interest. The method comprises the steps of (a) contacting the test sample with a binding partner that binds specifically to a marker (or target) for the virus, which if present in the sample binds to the binding partner to produce a marker-binding partner complex, and (b) detecting the presence or
10 absence of the complex. The presence of the complex is indicative of the presence of the virus in the test sample.

In one embodiment, the marker is an anti-US-type or anti-US-subtype antibody, for example, an immunoglobulin G (IgG) or an immunoglobulin M (IgM) molecule, present in the sample of interest, and the binding partner is an isolated polypeptide chain defining an epitope
15 that binds specifically to the marker. In such a case, it is contemplated that the test sample is a body fluid sample, for example, blood, serum or plasma, harvested from an individual under investigation. In a preferred embodiment, the polypeptide chain defining a US-type or US-subtype specific epitope is immobilized on a solid support. Thereafter, the immobilized polypeptide chain is combined with the sample under conditions that permit the marker
20 antibody, for example, an anti-US-type or anti-US-subtype hepatitis E virus specific antibody, present in the sample to bind to the immobilized polypeptide. Thereafter, the presence or absence of bound antibody can be detected using, for example, a second antibody or an antigen binding fragment thereof, for example, an anti-human antibody or an antigen binding fragment thereof, labeled with a detectable moiety.

25 It is contemplated that many different US-type and US-subtype specific polypeptides may be useful as a binding partner in the practice of this embodiment of the invention. For example, in one preferred embodiment of the invention, it is contemplated that the binding partner may be at least a portion, for example, at least 5, preferably at least 8, more preferably at least 15 and even more preferably at least about 25 amino acid residues, of a polypeptide

chain selected from the group consisting of SEQ ID NOS:91, 92 and 93, including naturally occurring variants thereof, and which represent a unique amino acid sequence when compared to the corresponding amino acid sequences of members of the Burmese and Mexican families. Similarly, it is contemplated that the binding partner may be a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NOS:173, 174, or 175. In another preferred embodiment of the invention, it is contemplated that the binding partner may be at least a portion, for example, at least 5, preferably at least 8, more preferably at least 15 and even more preferably at least about 25 amino acid residues, of a polypeptide chain selected from the group consisting of SEQ ID NOS:166, 167 and 168, including naturally occurring variants thereof, and which represent a unique amino acid sequence when compared to the corresponding amino acid sequences of members of the Burmese and Mexican families. Similarly, it is contemplated that the binding partner may be a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NOS:176, 223 or 224.

In another embodiment of the invention, the marker is a polypeptide chain unique for a member of the US-type or US-subtype families of HEV, and the binding partner preferably is an isolated antibody, for example, a polyclonal or monoclonal antibody, that binds to an epitope on the marker polypeptide chain. The binding partner may be either labeled with a detectable moiety or immobilized on a solid support. For example, it is contemplated that practice of this embodiment of the invention may be facilitated by immobilizing on a solid support, a first antibody that binds a first epitope on the marker polypeptide of interest. A test sample to be analyzed then is combined with the solid support under conditions that permit the immobilized antibody to bind the marker polypeptide. Thereafter, the presence or absence of bound marker polypeptide chain may be determined using, for example, a second antibody conjugated with a detectable moiety which binds to a second, different epitope on the marker polypeptide chain.

An antibody useful in the practice of this embodiment of the invention preferably is capable of binding specifically to a polypeptide chain selected from the group consisting of SEQ ID NOS:91, 92, and 93, including naturally occurring variants thereof, and has a higher binding affinity for such a polypeptide chain relative to the corresponding sequences of members of the Burmese and Mexican families. It is contemplated that an antibody useful in

the practice of the invention preferably is capable of binding specifically to a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NOS:173 or 175. This antibody being further characterized as, under similar conditions, preferably having a lower affinity for, and most preferably failing to bind the amino acid sequence set forth in SEQ. ID NOS:169 or 171 or to the regions in the Burmese and Mexican strains that correspond to SEQ ID NO:175. Similarly, it is contemplated that an antibody useful in the practice of the invention preferably is capable of binding specifically to a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NOS:174 or 176. This antibody being further characterized as, under similar conditions, preferably having a lower affinity for, and most preferably failing to bind the amino acid sequence set forth in SEQ. ID NOS:170 or 172 or to the regions in the Burmese and Mexican strains that correspond to SEQ ID NO:176.

Similarly, it is contemplated that an antibody useful in the practice of this embodiment of the invention preferably is capable of binding specifically to a polypeptide chain selected from the group consisting of SEQ ID NOS:166, 167, and 168, including naturally occurring variants thereof, and has a higher binding affinity for such a polypeptide chain relative to the corresponding sequences of members of the Burmese and Mexican families. It is contemplated that an antibody useful in the practice of the invention preferably is capable of binding specifically to a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NO: 223. This antibody being further characterized as, under similar conditions, preferably having a lower affinity for, and most preferably failing to bind the amino acid sequences set forth in SEQ. ID NOS:170 or 172. Similarly, it is contemplated that an antibody useful in the practice of the invention preferably is capable of binding specifically to a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NO:224. This antibody being further characterized as, under similar conditions, preferably having a lower affinity for, and most preferably failing to bind the amino acid sequence set forth in SEQ ID NOS:169 or 171.

In another embodiment of the invention, the marker is a nucleic acid sequence defining at least a portion of a genome of a US-type or US-subtype E virus, or a sequence complementary thereto. Similarly, it is contemplated that the binding partner is an isolated nucleic acid sequence, for example, a deoxyribonucleic acid (DNA), ribonucleic acid (RNA) or

peptidyl nucleic acid (PNA) sequence, preferably comprising 8-100 nucleotides, more preferably comprising 10 to 75 nucleotides and mostly preferably comprising 15-50 nucleotides, which is capable of hybridizing specifically, for example, under specific hybridization conditions or under specific PCR annealing conditions, to the nucleotide
5 sequence set forth in SEQ ID NOS:89 or 164.

Practice of this embodiment of the invention may be facilitated, for example, by isolating nucleic acids from the sample of interest. Thereafter, the resulting nucleic acids, may be fractionated by, for example, gel electrophoresis, transferred to, and immobilized onto a solid support, for example, nitrocellulose or nylon membrane, or alternatively may be
10 immobilized directly onto the solid support via conventional dot blot or slot blot methodologies. The immobilized nucleic acid then may be probed with a preselected nucleic acid sequence labeled with a detectable moiety, that hybridizes specifically to the marker sequence. Alternatively, the presence of marker nucleic acid in a sample may be determined by standard amplification based methodologies, for example, polymerase chain reaction (PCR)
15 wherein the production of a specific amplification product is indicative of the presence of marker nucleic acid in the sample.

In another aspect, the invention provides isolated US-type and US-subtype specific polypeptides sequences. These polypeptides include those described hereinabove in the section pertaining to US-type and US-subtype hepatitis E specific polypeptides chains useful as
20 binding partners. In a preferred embodiment, the isolated polypeptide chain comprises an amino acid sequence set forth in SEQ ID NO:93, SEQ ID NO:168, SEQ ID NO:173, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176, SEQ ID NO:223 or SEQ ID NO:224. It is contemplated that these and other US-type and US-subtype specific polypeptide chains may be employed in an assay format for detecting the presence of anti-US-type of US-subtype hepatitis
25 E specific antibodies in a sample. In addition, it is contemplated that these polypeptides may be used either alone or in combination with adjuvants for the production of antibodies in laboratory animals, or similarly, used in combination with pharmaceutically acceptable carriers as vaccines for either the prophylactic or therapeutic immunization of mammals.

In another aspect, the invention provides isolated anti-US-type or anti-US-subtype

hepatitis E specific antibodies, which include those discussed hereinabove in the section pertaining to antibodies useful as binding partners. In a preferred embodiment, the isolated antibody is capable of binding specifically to a polypeptide chain selected from the group consisting of a polypeptide encoded by an ORF 1 sequence of a US-type or a US-subtype hepatitis E virus, a polypeptide encoded by an ORF 2 sequence of a US-type or a US-subtype hepatitis E virus, or a polypeptide encoded by an ORF 3 sequence of a US-type or a US-subtype hepatitis E virus. In particular, it is contemplated that useful antibodies are characterized in that they are capable of binding specifically to a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NO:93, SEQ ID NO:168, SEQ ID NO:173, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176, SEQ ID NO:223 or SEQ ID NO:224. It is contemplated that these antibodies and other antibodies may be used to advantage in immunoassays for detecting the presence in a sample of members of the US-type or US-subtype hepatitis E families. The antibody may be used either in a direct immunoassay wherein the antibody itself preferably is labeled with a detectable moiety or in an indirect immunoassay wherein the antibody itself provides a target for a second binding partner, *e.g.*, a second antibody labeled with a detectable moiety. Furthermore, it is contemplated that these antibodies may be used in combination with, for example, a pharmaceutically acceptable carrier for use in the passive, therapeutic or prophylactic immunization of a mammal.

In another aspect, the invention provides isolated nucleic acid sequences such as those discussed in the previous section pertaining to the use of nucleic acids as a marker or a binding partner for detecting the presence of a US-type or US-subtype hepatitis E virus in a sample. In a preferred embodiment, the invention provides isolated nucleic acid sequences defining at least a portion of an ORF 1, ORF 2 or ORF 3 sequence of a US-type or US-subtype hepatitis E virus, or a sequence complementary thereto. It is contemplated that these and other nucleic acid sequences may be used, for example, as nucleotide probes and/or amplification primers for detecting the presence of a US-type or US-subtype hepatitis E virus in a sample of interest. In addition, it is contemplated the nucleic acid sequences or sequences complementary thereto may be combined with a pharmaceutically acceptable carrier for use in anti-sense therapy. Furthermore, it is contemplated the nucleic acid sequences may be integrated in vectors which may then be transformed or transfected into a host cell of interest. The host cells may then be

combined with a pharmaceutically acceptable carrier and used as a vaccine, for example, a recombinant vaccine, for immunizing a mammal, either prophylactically or therapeutically, against a preselected US-type or US-subtype hepatitis E virus.

The foregoing and other objects, features and advantages of the present invention will be made more apparent from the following detailed description of preferred embodiments of the invention.

Brief Description of the Drawings

The objects and features of the invention may be better understood by reference to the drawings described below in which,

Figure 1 is a schematic representation of a HEV genome showing the relative positions of the ORF 1, ORF 2, and ORF 3 regions.

Figure 2 is a graph showing levels of serum aspartate aminotransferase (boxes) and serum total bilirubin (diamonds) in patient USP-1 from day 1 of a hospital admission through day 37 post admission.

Figure 3 is a schematic representation of the HEV US-1 genome showing the relative positions of clones isolated during the course of this work.

Figure 4 is a schematic representation of the HEV US-2 genome showing the relative positions of clones isolated during the course of this work.

Figure 5 shows an unrooted phylogenetic tree depicting the relationship of nucleotide sequences from full length HEV US-1, HEV US-2, and 10 other HEV isolates. Branch lengths are proportional to the evolutionary distances between sequences. The scale representing nucleotide substitutions per position is shown. The internal node numbers indicate the bootstrap values (expressed as a percentage of all trees) obtained from 100 replicates. Isolates represented are Burmese, B1, B2; Chinese, C1, C2, C3, C4; Pakistan, P1; Indian, I1, I2; Mexican, M1; and United States, US-1, US-2.

Figure 6 shows an unrooted phylogenetic tree depicting the relationship of nucleotide sequences from the ORF 2/3 regions (*i.e.*, sequences corresponding to nucleotide residue numbers 5094-7114 of SEQ ID NO:89). Branch lengths are proportional to the evolutionary distances between sequences. The scale representing nucleotide substitutions per position is shown. The internal node numbers indicate the bootstrap values (expressed as a percentage of all trees) obtained from 100 replicates. Isolates represented are Burmese, B1, B2; Chinese, C1, C2, C3, C4; Pakistan, P1; Indian, I1, I2; Mexican, M1; Swine, S1; and United States, US-1, US-2.

Figure 7 is a graph showing levels of alanine aminotransferase (boxes), serum aspartate transferase (circles), and gamma-glutamyltransferase (triangles) in a macaque before and after inoculation with sera harvested from patient USP-2. Also shown are times when HEV US-2 RNA were present in serum and fecal samples, as well as times when anti-HEV US-2 IgM and IgG were detectable.

Figure 8 is a schematic representation of the Itl genome showing the relative positions of clones isolated during the course of this work.

Figures 9 shows alignments of Burmese (B1), Mexican (M1), Chinese (C1), Pakistan (P1) and US-1 showing the design of HEV consensus primers for ORF 1, ORF 2/3 and ORF 2. Preferred consensus primers are denoted by the highlighted boxes.

Figure 10 shows an unrooted phylogenetic tree depicting the relationship of ORF 1 nucleotide sequences 371 nucleotides in length and corresponding to residues 26-396 of SEQ ID NO:89. The scale representing nucleotide substitutions per position is shown. The internal node numbers indicate the bootstrap values (expressed as a percentage of all trees) obtained from 1000 replicates. Isolates represented are Burmese, B1, B2; Chinese, C1, C2, C3, C4; Pakistan, P1; Indian, I1, I2; Mexican, M1; Italian, It1; Greek, G1, G2; and United States, US-1, US-2.

Figure 11 shows an unrooted phylogenetic tree depicting the relationship of ORF 2 nucleotide sequences 148 nucleotides in length and corresponding to residues 6307-6454 of

SEQ ID NO:89. The scale representing nucleotide substitutions per position is shown. The internal node numbers indicate the bootstrap values (expressed as a percentage of all trees) obtained from 1000 replicates. Isolates represented are Burmese, B1, B2; Chinese, C1, C2, C3, C4; Pakistan, P1; Indian, I1, I2; Mexican, M1; Italian, It1; Greek, G1, G2; Swine, S1; and United States, US-1 and US-2.

Figure 12 shows a schematic representation of preferred HEV-US recombinant protein constructs. In 12A, the ORF 2 and ORF 3 structural proteins of HEV are shown with the first and last amino acid positions designated. The presence of immunodominant epitopes are indicated by lines within the ORFs. Figure 12B shows an ORF 3 region that was cloned into an expression vector, with the first and last amino acid positions designated (SEQ ID NO:203 or SEQ ID NO:204). Figure 12C shows an ORF 2 region that was cloned into an expression vector, with the first and last amino acid positions designated (SEQ ID NO:199 or 200). Figure 12D shows an ORF 3/2 chimeric construct cloned into an expression vector with the first and last amino acid positions of each component of the chimeric construct designated (SEQ ID NO:206 or 207). The sequence omitted from the ORF 3/2 construct is indicated with a dashed line. In Figures 12B-12D, the presence of a FLAG[®] peptide at the carboxyl terminus of each construct is indicated by a solid box.

Figure 13 is a graph showing levels of alanine aminotransferase (square), IgG (circle) and IgM (star) in a macaque before and after inoculation with sera harvested from patient USP-2.

Figure 14 shows an unrooted phylogenetic tree depicting the relationship of ORF 1 nucleotide sequences 371 nucleotides in length and corresponding to residues 26-396 of SEQ ID NO:89. The scale representing nucleotide substitutions per position is shown. The internal node numbers indicate the bootstrap values (expressed as a percentage of all trees) obtained from 1000 replicates. Isolates represented are Burmese, B1, B2; Chinese, C1, C2, C3, C4; Pakistan, P1; Indian, I1, I2; Mexican, M1; Italian, It1; Greek, G1, G2; Austrian, Au1; Argentine, Ar1, Ar2; and United States, US-1, US-2.

Figure 15 shows an unrooted phylogenetic tree depicting the relationship of ORF 2 nucleotide sequences 148 nucleotides in length and corresponding to residues 6307-6454 of SEQ ID NO:89. The scale representing nucleotide substitutions per position is shown. The internal node numbers indicate the bootstrap values (expressed as a percentage of all trees) obtained from 1000 replicates. Isolates represented are Burmese, B1, B2; Chinese, C1, C2, C3, C4; Pakistan, P1; Indian, I1, I2; Mexican, M1; Italian, It1; Greek, G1, G2; Austrian, Au1; Argentine, Ar2; Swine, S1; and United States, US-1 and US-2.

Figure 16 shows an unrooted phylogenetic tree depicting the relationship of ORF 2 nucleotide sequences 98 nucleotides in length and corresponding to residues 6354-6451 of SEQ ID NO:89. The scale representing nucleotide substitutions per position is shown. The internal node numbers indicate the bootstrap values (expressed as a percentage of all trees) obtained from 1000 replicates. Isolates represented are Burmese, B1, B2; Chinese, C1, C2, C3, C4; Pakistan, P1; Indian, I1, I2; Mexican, M1; Italian, It1; Greek, G1, G2; Austrian, Au1; Argentine, Ar1, Ar2; Swine, S1; and United States, US-1 and US-2.

Detailed Description of the Invention

As mentioned above, this invention is based, in part, upon the discovery of a new family of human hepatitis E viruses. The newly discovered family of hepatitis E viruses fall within a class referred to hereinafter as a US-type hepatitis E virus. Furthermore, as mentioned above, two members of the US-type family were identified in sera obtained from two individuals living in the United States of America. These two members together belong to a subclass of the US-type hepatitis E virus, referred to hereinafter as a US-subtype hepatitis E virus. The discovery of the US-type and US-subtype hepatitis E viruses enables the development of methods and compositions for detecting the presence of a US-type of US-subtype hepatitis E virus in individuals who heretofore have not been diagnosed as suffering from hepatitis based on commercially available hepatitis detection kits, as well as methods and compositions for immunizing an individual against such a virus.

In one aspect, the invention pertains to a method of detecting the presence of a US-type or US-subtype hepatitis E virus in a test sample. The method comprises the steps of (a)

contacting the sample with a binding partner that binds specifically to a marker for such a virus, which if present in the sample binds to the binding partner to produce a marker-binding protein complex, and (b) detecting the presence or absence of the complex. The presence of the complex is indicative of the presence of the virus in the sample. Based on the discovery of the US-type and US-subtype hepatitis E virus disclosed herein, it will be apparent that a variety of assays, for example, protein- or nucleic acid-based assays, may be produced for detecting the presence of the virus in a sample. Protein-based assays may include, for example, conventional immunoassays, and nucleic acid-based assays may include, for example, conventional probe hybridization or nucleic acid sequence amplification assays, all of which are well known and thoroughly discussed in the art.

In another aspect, the invention provides reagents, for example, antibodies, epitope containing polypeptide chains, and nucleotide sequences that may be used to develop vaccines for immunizing, either prophylactically or therapeutically, an individual against a US-type or US-subtype hepatitis E virus.

I. Definitions

So that the invention may be more readily understood, certain terms as used herein are defined hereinbelow.

As used herein, the term "US-type" hepatitis E virus is understood to mean any human virus (*i.e.*, capable of infecting a human) that is serologically distinct from hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis D virus (HDV) and hepatitis G virus (HGV) and comprising a single stranded RNA genome defining at least one open reading frame and having a nucleotide sequence greater than 79.7% identity to the nucleotide sequence defined by residues 6307-6454 of SEQ ID NO:89.

As used herein, the term "US-subtype" hepatitis E is understood to mean any human virus (*i.e.*, capable of infecting a human) that is serologically distinct from hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis D virus (HDV) and hepatitis G virus (HGV) and comprising a single stranded RNA genome defining at least one

open read frame and having a nucleotide sequence greater than 90.5% identity to the nucleotide sequence defined by residues 6307-6454 of SEQ ID NO:89.

As used herein, the term, "test sample" is understood to mean any sample, for example, a biological sample, which contains the marker (for example, an antibody, antigenic protein or peptide, or nucleotide sequence) to be tested. Preferred test samples include tissue or body fluid samples isolatable from an individual under investigation. Preferred body fluid samples include, for example, blood, serum, plasma, saliva, sputum, semen, urine, feces, bile, spinal fluid, breast exude, ascities, and peritoneal fluid. Another preferred test sample is a cell line and more preferably, a mammalian cell line. A most preferred cell line is a human fetal kidney cell line.

As used herein, the term "open reading frame" or "ORF" is understood to mean a region of a polynucleotide sequence capable of encoding one or more polypeptide chains. The region may represent an entire coding sequence, *i.e.*, beginning with an initiation codon (*e.g.*, ATG (AUG)) and ending at a termination codon (*e.g.*, TAA (UAA), TAG (UAG), or TGA (UGA)), or a portion thereof.

As used herein, the term "polypeptide chain" is understood to mean any molecular chain of amino acids and does not refer to a specific length of the product. Thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide chain.

As used herein, the term "epitope", as used synonymously with "antigenic determinant", is understood to mean at least a portion of an antigen capable of being specifically bound (*i.e.*, bound with an affinity greater than about 10^5 M^{-1} , and more preferably with an affinity greater than about 10^7 M^{-1}) by an antibody variable region. Conceivably, an epitope may comprise three amino acids in a spatial conformation unique to the epitope. Generally, an epitope comprises at least five amino acids, and more usually, at least eight to ten amino acids. Methods of examining spatial conformation are known in the art and include, for example, x-ray crystallography and two-dimensional nuclear magnetic resonance.

A polypeptide is "immunologically reactive" with an antibody when it binds to an antibody due to antibody recognition of a specific epitope defined by the polypeptide chain. Immunological reactivity may be determined by antibody binding, more particularly by the kinetics of antibody binding, and/or by a competitive binding study. If a preselected antibody is immunologically reactive with a first antigen but is not immunologically reactive or is less immunologically reactive with a second, different antigen, then the two antigens are considered to be serologically distinct. As used herein, the term "affinity" is understood to mean a measure of reversible interaction between two molecules (for example, between an antibody and an antigen). The higher the affinity, the stronger the interaction between the two molecules.

As used herein, the term "detectable moiety" is understood to mean any signal generating compound, for example, chromogen, a catalyst such as an enzyme, a luminescent compound such as dioxetane, acridinium, phenanthridinium and luminol, a radioactive element, and a visually detectable label. Examples of enzymes include alkaline phosphatase, horseradish peroxidase, beta-galactosidase, and the like. Although the selection of a particular detectable moiety is not critical, the detectable moiety will be capable of producing a signal either by itself or in conjunction with one or more additional substances.

As used herein, the term "solid support" is understood to mean any plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface. Useful surfaces include, for example, the surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cell, or duracyte. Suitable solid supports are not critical to the practice of the invention and can be selected by one skilled in the art. Suitable methods for immobilizing peptides on solid phases include ionic, hydrophobic, covalent interactions and the like. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid support can retain an additional receptor which has the ability to attract and immobilize the capture reagent.

It is contemplated that the solid support also may comprise any suitable porous material with sufficient porosity to allow access by detection antibodies and a suitable surface affinity to bind antigens. Microporous structures generally are preferred, but materials with gel structure

in the hydrated state may be used as well. All of these materials may be used in suitable shapes, such as films, sheets, or plates, or they may be coated onto or bonded or laminated to appropriate inert carriers, such as paper, glass, plastic films, or fabrics.

Other embodiments which utilize various other solid supports also are contemplated and are within the scope of this invention. For example, ion capture procedures for immobilizing an immobilizable reaction complex with a negatively charged polymer, described in EP Publication No. 0 326 100 and EP Publication No. 0 406 473, can be employed according to the present invention to effect a fast solution-phase immunochemical reaction. An immobilizable immune complex is separated from the rest of the reaction mixture by ionic interactions between the negatively charged poly-anion/immune complex and the previously treated, positively charged porous matrix and detected by using various signal generating systems previously described, including those described in chemiluminescent signal measurements as described in EP Publication No. 0 273 115.

Also, the methods of the present invention can be adapted for use in systems which utilize microparticle technology including automated and semi-automated systems wherein the solid phase comprises a microparticle (magnetic or non-magnetic). Such systems include those described in U.S. Patent Nos. 5,089,424 and 5,244,630, issued February 18, 1992 and September 14, 1993, respectively.

The use of scanning probe microscopy (SPM) for immunoassays also is a technology to which the monoclonal antibodies of the present invention are easily adaptable. In scanning probe microscopy, in particular in atomic force microscopy, the capture phase, for example, at least one of the monoclonal antibodies of the invention, is adhered to a solid phase and a scanning probe microscope is utilized to detect antigen/antibody complexes which may be present on the surface of the solid phase. The use of scanning tunneling microscopy eliminates the need for labels which normally must be utilized in many immunoassay systems to detect antigen/antibody complexes. The use of SPM to monitor specific binding reactions can occur in many ways. In one embodiment, one member of a specific binding partner (analyte specific substance which is the monoclonal antibody of the invention) is attached to a surface suitable for scanning. The attachment of the analyte specific substance may be by adsorption to a test

piece which comprises a solid phase of a plastic or metal surface, following methods known to those of ordinary skill in the art. Or, covalent attachment of a specific binding partner (analyte specific substance) to a test piece which test piece comprises a solid phase of derivatized plastic, metal, silicon, or glass may be utilized. Covalent attachment methods are known to those skilled in the art and include a variety of means to irreversibly link specific binding partners to the test piece. If the test piece is silicon or glass, the surface must be activated prior to attaching the specific binding partner. Also, polyelectrolyte interactions may be used to immobilize a specific binding partner on a surface of a test piece by using techniques and chemistries described in EP Publication No. 0 322 100 and EP Publication No. 0 406 473. The preferred method of attachment is by covalent attachment. Following attachment of a specific binding member, the surface may be further treated with materials such as serum, proteins, or other blocking agents to minimize non-specific binding. The surface also may be scanned either at the site of manufacture or point of use to verify its suitability for assay purposes. The scanning process is not anticipated to alter the specific binding properties of the test piece.

As used herein, the terms "nucleotide sequence" or "nucleic acid sequence" is understood to mean any polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. The term refers to the primary structure of the molecule. Thus, the term includes double- and single-stranded DNA, as well as double- and single-stranded RNA. It also includes modifications, for example, by methylation and/or by capping, and unmodified forms of the polynucleotide.

As used herein, the term "primer" is understood to mean a specific oligonucleotide sequence complementary to a target nucleotide sequence which is capable of hybridizing to the target nucleotide sequence and serving as an initiation point for nucleotide polymerization catalyzed by DNA polymerase, RNA polymerase or reverse transcriptase.

When referring to a nucleic acid fragment, such a fragment is considered to "specifically hybridize" or to "specifically bind" to an HEV US-type or US-subtype polynucleotide or variants thereof, if, within the linear range of detection, the hybridization results in a stronger signal relative to the signal that would result from hybridization to an equal amount of a polynucleotide from other than an HEV US-type, US- subtype or variant thereof. A signal

which is "stronger" than another is one which is measurable over the other by the particular method of detection.

Also, when referring to a nucleic acid fragment, such a fragment is considered to hybridize under specific hybridization conditions if it specifically hybridizes under (i) typical hybridization and wash conditions, such as those described, for example, in Maniatis, (1st Edition, pages 387-389, 1982) where preferred hybridization conditions are those of lesser stringency and more preferred, higher stringency; or (ii) standard PCR conditions (Saiki, R.K. *et al.*) or "touch-down" PCR conditions (Roux, K.H., (1994), *Biotechniques*, 16:812-814).

As used herein, the term "probe" is understood to mean any nucleotide or nucleotide analog (*e.g.*, PNA) containing a sequence which can be used to identify specific DNA or RNA present in samples bearing the complementary sequence.

As used herein, the term "PNA" is used to mean peptide nucleic acid analog which may be utilized in a procedure such as an assay described herein to determine the presence of a target. "MA" denotes a "morpholino analog" which may be utilized in a procedure such as an assay described herein to determine the presence of a target. See, for example, U.S. Patent No. 5,378,841, which is incorporated herein by reference. PNAs typically are neutrally charged moieties which can be directed against RNA targets or DNA. PNA probes used in assays in place of, for example, the DNA probes of the present invention, offer advantages not achievable when DNA probes are used. These advantages include manufacturability, large scale labeling, reproducibility, stability, insensitivity to changes in ionic strength and resistance to enzymatic degradation which is present in methods utilizing DNA or RNA. These PNAs can be labeled with such signal generating compounds as fluorescein, radionucleotides, chemiluminescent compounds, and the like. PNAs or other nucleic acid analogs such as MAs thus can be used in assay methods in place of DNA or RNA. Although assays are described herein utilizing DNA probes, it is within the scope of the routine that PNAs or MAs can be substituted for RNA or DNA with appropriate changes if and as needed in assay reagents.

When referring to a nucleic acid fragment, such a fragment is considered to "specifically hybridize" or to "specifically bind" to an HEV US-type or US-subtype polynucleotide or

variants thereof, if, within the linear range of detection, the hybridization results in a stronger signal relative to the signal that would result from hybridization to an equal amount of a polynucleotide from other than an HEV US-type, US- subtype or variant thereof. A signal which is "stronger" than another is one which is measurable over the other by the particular method of detection.

Also, when referring to a nucleic acid fragment, such a fragment is considered to hybridize under specific hybridization conditions if it specifically hybridizes under (i) typical hybridization and wash conditions, such as those described, for example, in Maniatis, (1st Edition, pages 387-389, 1982) where preferred hybridization conditions are those of lesser stringency and more preferred, higher stringency; or (ii) standard PCR conditions (Saiki, R.K. *et al.*) or "touch-down" PCR conditions (Roux, K.H., (1994), *Biotechniques*, 16:812-814).

II. Detection Methods and Reagents

It is contemplated that the detection methods of the invention may employ a variety of protein-based or nucleic acid-based assays which are described in detail below.

It is contemplated that a reagent for the detection of virus or markers thereof may be either an anti-US-type and/or US-subtype hepatitis E virus antibody, a US-type and/or US-subtype specific polypeptide, or a nucleic acid defining at least a portion of the genome of a US-type and/or US-subtype hepatitis E virus or a nucleic acid sequence complementary thereto.

II. (i) Protein-based Assays

A. Marker Antibodies: It is contemplated that if the viral marker is an anti-US-type or anti-US-subtype specific antibody, for example, an IgG or an IgM, molecule circulating in the blood stream of an individual of interest, the binding partner preferably is a polypeptide defining an epitope that binds specifically to the marker.

In a preferred protocol for detecting the presence of anti-US-type or anti-US-subtype hepatitis E virus antibodies in a test sample, the protocol preferably comprises the following

steps which include: (a) providing an antigen comprising an immunologically reactive US-type or US-subtype specific polypeptide chain comprising at least 5, more preferably at least 8, even more preferably at least 15, and most preferably at least 25 contiguous amino acid residues and bindable by the antibody; (b) incubating the antigen with the test sample under conditions that permit formation of an antibody-antigen complex; and (c) detecting the presence of the complex.

It is contemplated that many, different US-type or US-subtype specific polypeptides may be useful as a binding partner for the detection of anti-US-type or anti-US-subtype antibodies. For example, it is contemplated that the polypeptide chain may be an amino acid sequence defined by SEQ ID NOS:91, 92 or 93 or an immunologically reactive fragment thereof containing, preferably at least 5, more preferably at least 8, even more preferably at least 15, and most preferably at least about 25 contiguous amino acid residues, of the polypeptide chain set forth in SEQ ID NOS:91, 92, or 93, and which represent a unique amino acid sequence when compared to the corresponding amino acid sequences of members of the Burmese and Mexican families. The Burmese family *i.e.*, "Burmese-like" strains, as used herein, presently comprises strains referred to herein as B1, B2, I1, I2, C1, C2, C3, C4 and P1 and the Mexican family presently comprises strain M1.

It is contemplated that the binding partner may be a polypeptide selected from the group consisting of polypeptides defined by SEQ ID NOS:91, 92, and 93, including naturally occurring variants thereof. As used herein the term "naturally occurring variants thereof" with respect to the polypeptide defined by SEQ ID NO:91 is understood to mean any amino acid sequence that is at least 84%, preferably at least 86%, more preferably at least 89% and even more preferably at least 95% identical to residues 1 through 1698 of SEQ ID NO:91. As used herein the term "naturally occurring variants thereof" with respect to the polypeptide defined by SEQ ID NO:92 is understood to mean any amino acid sequence that is at least 93%, preferably at least 95%, and even more preferably at least 98% identical to residues 1 through 660 of SEQ ID NO:92. As used herein the term "naturally occurring variants thereof" with respect to the polypeptide defined by SEQ ID NO:93 is understood to mean any amino acid sequence that is

at least 85.4%, preferably at least 87.4%, more preferably at least 90.4% and even more preferably at least 95% identical to residues 1 through 122 of SEQ ID NO:93.

Furthermore, it is contemplated that the binding partner may be a polypeptide encoded by a portion of an ORF 1 sequence. Proteins encoded by the ORF 1 sequence include, for example, a methyltransferase protein, a protease, a Y domain protein, an X domain protein, a helicase protein, a hypervariable region protein, and an RNA-dependent RNA polymerase protein. Accordingly, it is contemplated that a useful methyltransferase protein preferably has at least 92.3%, more preferably has at least 94.3%, and most preferably has at least 97.3% identity to residues 1-231 of SEQ ID NO:91. Also, it is contemplated that a useful protease protein preferably has at least 70.3%, more preferably has at least 72.3%, and most preferably has at least 75.3% identity to residues 424-697 of SEQ ID NO:91. Also, it is contemplated that a useful Y domain protein preferably has at least 94.6%, more preferably has at least 96.6% and most preferably has at least 99.6% identity to residues 207-424 of SEQ ID NO:91. Also it is contemplated that a useful X domain protein preferably has at least 83.4%, more preferably has at least 85.4% and most preferably has at least 88.4% identity to residues 789-947 of SEQ ID NO:91. Also, it is contemplated that a useful helicase protein has at least 92%, more preferably has at least 94% and most preferably at least 93% identity to residues 965-1197 of SEQ ID NO:91. Also, it is contemplated that a useful hypervariable region protein has at least 28.7%, more preferably has at least 30.7%, and most preferably has at least 33.7% identity to the residues 698-788 of SEQ ID NO:91. Also, it is contemplated that a useful RNA-dependent RNA polymerase has at least 88.8%, more preferably has at least 90.8%, and most preferably has at least about 93.8% identity to residues 1212-1698 of SEQ ID NO:91.

Furthermore, it is contemplated that the binding partner may be a polypeptide chain having an amino acid sequence defined by SEQ ID NOS:166, 167 or 168, or an immunologically reactive fragment thereof containing 5, preferably at least 8, more preferably at least 15 and most preferably at least 25 contiguous amino acid residues of the polypeptide chain set forth in SEQ ID NOS:166, 167 or 168, and which represent a unique amino acid sequence when compared to the corresponding amino acid sequences of members of the Burmese and Mexican families. Similarly, it is contemplated that the binding partner may be a

polypeptide selected from the group consisting of SEQ ID NOS:166, 167 and 168, including naturally occurring variants thereof. As used herein, the term “naturally occurring variants thereof” with respect to the polypeptide defined by SEQ ID NO:166 is understood to mean any amino acid sequence that is at least 83.9%, preferably at least 85.9%, more preferably at least 88.9%, and most preferably at least 95% identical to residues 1 through 1708 of SEQ ID NO:166. As used herein, the term “naturally occurring variants thereof” with respect to the polypeptide defined by SEQ ID NO:167 is understood to mean any amino acid sequence that is at least 93%, preferably at least 95%, and most preferably at least 98% identical to residues 1 through 660 of SEQ ID NO:167. As used herein, the term “naturally occurring variants thereof” with respect to the polypeptide defined by SEQ ID NO:168 is understood to mean any amino acid sequence that is at least 85.4%, preferably at least 87.4%, more preferably at least 90.4%, and even more preferably at least 95% identical to residues 1 through 122 of SEQ ID NO:168.

Furthermore, it is contemplated that the binding partner may be a polypeptide encoded by a portion of the HEV US-2 ORF 1, including, for example, a methyltransferase protein, a protease, a Y domain protein, an X domain protein, a helicase protein, a hypervariable region protein and an RNA-dependent RNA polymerase protein, or a variant thereof. Accordingly, it is contemplated that a useful methyltransferase protein preferably has at least 92.7%, more preferably has at least 94.7%, and most preferably has at least 97.7% identity to residues 1-240 of SEQ ID NO:166. Also, it is contemplated that a useful protease protein preferably has at least 69.6%, more preferably has at least 71.6%, and most preferably has at least 74.6% identity to residues 433-706 of SEQ ID NO:166. Also, it is contemplated that a useful Y domain protein preferably has at least 94.6%, more preferably has at least 96.6%, and most preferably has at least 99.6% identity to residues 216-433 of SEQ ID NO:166. Also it is contemplated that a useful X domain protein preferably has at least 82.8%, more preferably has at least 84.8%, and most preferably has at least 87.8% identity to residues 799-957 of SEQ ID NO:166. Also, it is contemplated that a useful helicase protein has at least 92.8%, more preferably has at least 94.8%, and most preferably has at least 97.8% identity to residues 975-1207 of SEQ ID NO:166. Also, it is contemplated that a useful hypervariable region protein has at least 27%, more preferably has at least 29%, and most preferably has at least 31%

identity to the residues 707-798 of SEQ ID NO:166. Also, it is contemplated that a useful RNA-dependent RNA polymerase has at least 88.7%, more preferably has at least 90.7%, and most preferably has at least 93.7% identity to residues 1222-1708 of SEQ ID NO:166.

With regard to the identification of US-type or US-subtype specific epitopes, it is contemplated that one skilled in the art in possession of nucleic acid sequences defining and/or amino acid sequences encoded by at least a portion of the genome of a US-type or US-subtype hepatitis E virus can map potential epitope sites using conventional technologies well known and thoroughly discussed in the art. In addition to the use of commercially available software packages which identify potential epitope sites in a given sequence, it is possible to identify potential epitopes by comparison of amino acid sequences encoded by such a genome with sequences encoded by the genomes of other strains of HEV whose antigenic sites have already been elucidated. See, for example, U.S. Patent Nos: 5,686,239, 5,741,490 and 5,770,689. Epitopes currently identified are shown in Figure 1, and include epitopes referred to in the art as 8-5 (SEQ ID NOS:93 AND 168), 4-2 (position 90-122 of SEQ ID NOS:93 and 168), SG3 (SEQ ID NOS:175 AND 176), 3-2 (position 613-654 of SEQ ID NOS:92 and 167) and 3-2e (position 613-660 of SEQ ID NOS:92 and 167). A method for calculating antigenic index is described by Jameson and Wolf (CABIOS, 4(1), 181-186 [1988]).

For example, two epitopes of interest are discussed in detail below and are referred to as 3-2e and 4-2 which are encoded by portions of ORF 2 and ORF 3 of the hepatitis E genome, respectively. These epitopes were identified in the Burmese strains of HEV (referred to below as B 3-2e (SEQ ID NO:172) and B 4-2 (SEQ ID NO:171)), and in the Mexican strain of HEV (referred to below as M 3-2e (SEQ ID NO:170) and M 4-2 (SEQ ID NO:169)). Similar epitopes were identified in HEV US-1 based on amino acid sequence comparisons, and are referred to below as U3-2e (SEQ ID NO:174) and U4-2 (SEQ ID NO:173). Similar epitopes were identified in HEV US-2, also based on amino acid sequence comparisons, and are referred to below as US-2 3-2e (SEQ ID NO:223) and US-2 4-2 (SEQ ID NO:224).

In addition, potential epitopes may be identified using screening procedures well known and thoroughly documented in the art. For example, based on the nucleic acid sequences defining either the entire or portions of the HEV US-1 or the HEV US-2 genome, it is possible

to generate an expression library, which, after expression can be screened to identify epitopes. For example, nucleic acid fragments representative of the HEV US-1 or the HEV US-2 genome can be cloned into the lambda-gt11 expression vector to produce a lambda-gt11 library, for example, a cDNA library. The library then is screened for encoded epitopes that can bind specifically with sera derived from individuals identified as being infected with HEV US-1 or HEV US-2. See, for example, Glover (1985) in "DNA Cloning Techniques, A Practical Approach", IRL Press, pp. 49-78. Typically, about 10^6 - 10^7 phage are screened, from which positive phage are identified, purified, and then tested for specificity of binding to sera from different individuals previously infected with HEV US-1 or HEV US-2. Phage which bind selectively to antibodies present in sera or plasma from the individual are selected for additional characterization. Once identified, an amino acid sequence of interest may be produced in large scale either by use of conventional recombinant DNA methodologies or by conventional peptide synthesis methodologies, well known and thoroughly documented in the art.

b. Marker Polypeptides: It is contemplated that if the marker is a US-type or US-subtype virus or a specific polypeptide thereof, the binding partner useful in the practice of the invention preferably is an antibody, for example, a polyclonal or monoclonal antibody, that binds to an epitope on the virus or marker polypeptide. The binding partner may be either labeled with a detectable moiety or immobilized on a solid support. In particular, the antibodies useful in the practice of this embodiment preferably are capable of binding specifically to a US-type or US-subtype specific polypeptide chain preferably at least 5, more preferably at least 8, even more preferably at least 15, and most preferably at least 25 contiguous amino acid residues in length which is unique with respect to the corresponding amino acid sequence found in members of the Burmese and Mexican families.

An antibody useful in the practice of this embodiment of the invention preferably is capable of binding specifically to a polypeptide chain selected from the group consisting of SEQ ID NOS:91, 92, and 93, including naturally occurring variants thereof, and has a higher binding affinity for such a polypeptide chain relative to the corresponding sequences of members of the Burmese and Mexican families. It is contemplated that an antibody useful in the practice of the invention preferably is capable of binding specifically to a polypeptide chain

comprising the amino acid sequence set forth in SEQ ID NO:173 or 175. This antibody being further characterized as, under similar conditions, preferably having a lower affinity for, and most preferably failing to bind the amino acid sequence set forth in SEQ. ID NOS:169 or 171 or regions in the Burmese and Mexican strains that correspond to SEQ ID NO:175. Similarly, it is contemplated that an antibody useful in the practice of the invention preferably is capable of binding specifically to a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NOS:174 or 176. This antibody being further characterized as, under similar conditions, preferably having a lower affinity for, and most preferably failing to bind the amino acid sequence set forth in SEQ ID NOS:170 or 172 or regions in the Burmese and Mexican strains that correspond to SEQ ID NO:176.

Similarly, it is contemplated that an antibody useful in the practice of this embodiment of the invention preferably is capable of binding specifically to a polypeptide chain selected from the group consisting of SEQ ID NOS:166, 177, and 168, including naturally occurring variants thereof, and has a higher binding affinity for such a polypeptide chain relative to the corresponding sequences of members of the Burmese and Mexican families. It is contemplated that an antibody useful in the practice of the invention preferably is capable of binding specifically to a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NO:223. This antibody being further characterized as, under similar conditions, preferably having a lower affinity for, and most preferably failing to bind the amino acid sequences set forth in SEQ. ID NOS:170 or 172. Similarly, it is contemplated that an antibody useful in the practice of the invention preferably is capable of binding specifically to a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NO:224. This antibody being further characterized as, under similar conditions, preferably having a lower affinity for, and most preferably failing to bind the amino acid sequence set forth in SEQ ID NOS:169 or 171.

The antibodies or antigen binding fragments thereof as described herein can be provided individually to detect US-type or US-subtype specific antigens. Combinations of the antibodies (and antigen binding fragments thereof) provided herein also may be used together as components in a mixture or "cocktail" of at least two antibodies, both having different binding specificities to separate US-type or US-subtype specific antigens.

c. Antibody Production: It is contemplated that one skilled in the art, in possession of the nucleic acid sequences defining, or amino acid sequences encoded by at least a portion of the ORF 1, ORF 2 and/or ORF 3 sequences of a US-type or a US-subtype hepatitis E virus may be able to produce specific antibodies using techniques well known and thoroughly documented in the art. See, for example, Practical Immunology, Butt, N.R., ed., Marcel Dekker, NY, 1984. Briefly, an isolated target protein is used to raise antibodies in a xenogenic host, such as a mouse, pig, goat or other suitable mammal. Preferred antibodies are antibodies that bind specifically to an epitope on the target protein, preferably having a binding affinity greater than 10^5M^{-1} , and most preferably having a binding affinity greater than 10^7M^{-1} for that epitope. Typically, the target protein is combined with a suitable adjuvant capable of enhancing antibody production in the host, and injected into the host, for example, by intraperitoneal administration. Any adjuvant suitable for stimulating the host's immune response may be used to advantage. A commonly used adjuvant is Freund's complete adjuvant (an emulsion comprising killed and dried microbial cells, *e.g.*, from Calbiochem Corp., San Diego, CA or Gibco, Grand Island, NY). Where multiple antigen injections are desired, the subsequent injections comprise the antigen in combination with an incomplete adjuvant (*e.g.*, cell-free emulsion).

Polyclonal antibodies may be isolated from the antibody-producing host by extracting serum containing antibodies to the protein of interest. Monoclonal antibodies may be produced by isolating host cells that produce the desired antibody, fusing these cells with myeloma cells using standard procedures known in the immunology art (See for example, Kohler and Milstein, Nature (1975) 256:495), and screening for hybrid cells (hybridomas) that react specifically with the target protein and have the desired binding affinity.

In addition, it is contemplated that when small peptides are used their immunogenicity may be enhanced by coupling to solid supports. For example, an epitope or antigenic region or fragment of a polypeptide generally is relatively small, and may comprise about 8 to 10 amino acids or less in length. Fragments of as few as 3 amino acids may characterize an antigenic region. These polypeptides may be linked to a suitable carrier molecule when the polypeptide of interest provided folds to provide the correct epitope but yet is too small to be antigenic.

Preferred linking reagents and methodologies for their use are well known in the art and may include, without limitation, N-succinimidyl-3-(2-pyridylthio)propionate (SPDP) and succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC). Furthermore, polypeptides lacking sulfhydryl groups can be modified by adding a cysteine residue. These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino group on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. Other bifunctional coupling agents form a thioester rather than a disulfide linkage. Many of these thioether-forming agents are commercially available and are known to those of ordinary skill in the art. The carboxyl groups can be activated by combining them with succinimide or 1-hydroxyl-2-nitro-4-sulfonic acid, sodium salt. Any carrier which does not itself induce the production of antibodies harmful to the host can be used. Suitable carriers include proteins, polysaccharides such as latex functionalized sepharose, agarose, cellulose, cellulose beads, polymeric amino acids such as polyglutamic acid, polylysine, and no acid copolymers and inactive virus particles, among others. Examples of protein substrates include serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and yet other proteins known to those skilled in the art.

In addition, it is contemplated that biosynthetically produced antibody binding domains wherein the amino acid sequence of the binding domain is manipulated to enhance binding affinity to a preferred epitope also may be useful in the practice of the invention. A detailed description of their preparation can be found, for example, in Practical Immunology, Butt, W.R., ed., Marcel Dekker, New York, 1984. Optionally, a monovalent antibody fragment such as an Fab or an Fab' fragment may be utilized. Additionally, genetically engineered biosynthetic antibody binding sites may be utilized which comprise either 1) non-covalently associated or disulfide bonded synthetic V_H and V_L dimers, 2) covalently linked V_H - V_L single chain binding sites, 3) individual V_H or V_L domains, or 4) single chain antibody binding sites, as disclosed, for example, in U.S. Patent Nos. 5,091,513 and 5,132,405.

It is contemplated that intact antibodies (for example, monoclonal or polyclonal antibodies), antibody fragments or biosynthetic antibody binding sites that bind a US-type or

US-subtype hepatitis E virus specific epitope, will be useful in diagnostic and prognostic applications, and also, will be useful in passive immunotherapy.

d. Assay Formats: It is contemplated that both polypeptides which react immunologically with serum containing anti-US-type or anti-US-subtype hepatitis E virus specific antibodies, or antibodies raised against US-type or US-subtype hepatitis E specific epitopes will be useful in immunoassays to detect the presence of such a virus in a test sample of interest. Furthermore, it is contemplated that the presence of US-type or US-subtype hepatitis E virus in a sample may be detected using any of a wide range of immunoassay techniques, for example, direct assays, sandwich assays, and/or competition assays, currently known and thoroughly documented in the art. A variety of preferred assay formats are described in more detail below.

In one preferred format, the assay employs a sandwich format. Sandwich immunoassays typically are highly specific and very sensitive, provided that labels with good limits of detection are used. A detailed review of immunological assay design, theory and protocols can be found in numerous texts in the art, including Practical Immunology, Butt, W.R., ed., Marcell Dekker, New York, 1984.

In one type of sandwich format, a polypeptide (binding partner) which has been immobilized onto a solid support and is immunologically reactive with an anti-US-type or anti-US-subtype hepatitis E virus antibody (marker), is contacted with a test sample from an individual suspected of having been infected with the US-type or US-subtype hepatitis E virus, to form a mixture. The mixture then is incubated for a time and under conditions sufficient to form polypeptide/antibody complexes. Then, an indicator reagent comprising a monoclonal or a polyclonal antibody or a fragment thereof, which specifically binds to the test sample antibody, and labeled with a detectable moiety, is contacted with the antigen/antibody complexes to form a second mixture. The second mixture then is incubated for a time and under conditions sufficient to form antigen/antibody/antibody complexes. The presence of anti-US-type or anti-US-subtype hepatitis E antibody, if any, in the test sample is determined by detecting the presence of detectable moiety immobilized to the solid support. The amount of antibody present in the test sample is proportional to the signal generated. The use of biotin

and antibiotin, biotin and avidin, biotin and streptavidin, and the like, may be used to enhance the generated signal in the assay systems described herein.

In an alternative format of the above-described assay, the immunologically reactive polypeptide may be immobilized "indirectly" to the solid support, i.e. through a monoclonal or polyclonal antibody or fragment thereof which specifically binds that polypeptide. Alternatively, in another format, the assay components may be used in the reverse configuration, such that an antibody or antigen binding fragment thereof, which specifically binds the test sample antibody, *i.e.*, marker antibody (for example, IgG or IgM) and immobilized on the solid support is contacted with the test sample, for a time and under conditions sufficient to permit formation of antibody/antibody complexes. Then, an indicator reagent, for example, a US-type or US-subtype hepatitis E polypeptide immunologically reactive with captured test sample antibody and labeled with a detectable moiety, is incubated with the antibody/antibody complexes to form a second mixture for a time and under conditions sufficient to permit formation of antibody/antibody/antigen complexes. As above, the presence of antibody in the test sample, if any, that is captured by the capture antibody or antigen binding fragment thereof immobilized on the solid support is determined by detecting the measurable signal generated by the detectable moiety.

It is contemplated that the aforementioned sandwich assays also may be used to test for the presence of a US-type or US-subtype hepatitis E virus, or immunologically reactive polypeptides thereof in a test sample by routine modification of the above-described assay configurations. It is contemplated that such modifications would be well known to one skilled in the art.

In addition to the aforementioned sandwich assays, it is contemplated that competitive assays may also be employed in the practice of the invention. In this format, one or a combination of at least two antibodies, preferably monoclonal antibodies, which specifically bind to a US-type or US-subtype hepatitis E specific polypeptide chain can be employed as a competitive probe for the detection of antibodies to the US-type or the US-subtype specific protein. For example, a first HEV US-1 specific polypeptide chain such as one of the polypeptides disclosed herein, acting as a binding partner for the marker, is immobilized on a

solid support. A test sample suspected of containing antibody to HEV US-1 antigen then is incubated with the solid support together with an indicator reagent comprising, for example, an isolated anti-US-type or anti-US-subtype antibody that binds the immobilized HEV US-1 specific polypeptide chain and labeled with a detectable moiety, for a time and under conditions sufficient to form antigen/antibody complexes immobilized to the solid support. If the marker antibody is present in the test sample, then the marker antibody competes with the labeled indicator reagent for binding the immobilized polypeptide. As the amount of marker antibody present in the test sample increases, the amount of labeled indicator reagent that binds the immobilized polypeptide decreases. A reduction in the amount of indicator reagent bound to the solid phase can be quantitated. A measurable reduction in signal compared to the signal generated from a confirmed negative non-A, non-B, non-C, non-D, non-E hepatitis test sample also is indicative of the presence of anti-HEV US-1 antibody in the test sample. It is contemplated that similar protocols may be used to identify the presence in a test sample of other hepatitis E viruses falling within the US-type or US-subtype classes.

In yet another detection method, the antibodies of the present invention may be employed to detect the presence of US-type or US-subtype hepatitis E specific antigens in fixed tissue sections, as well as fixed cells by immunohistochemical analysis. Cytochemical analysis wherein these antibodies are labeled directly with a detectable moiety (*e.g.*, fluorescein, colloidal gold, horseradish peroxidase, alkaline phosphatase, etc.) or are labeled indirectly, for example, by means of a secondary antibody labeled with a detectable moiety also may be used in the practice of the invention.

In another assay format, the presence of antibody and/or antigen can be detected by means of a simultaneous assay, for example, as described in EP Publication No. 0 473 065. For example, a test sample is contacted simultaneously with (i) a capture reagent of a first analyte, wherein the capture reagent comprises a first binding member specific for a first analyte immobilized on a solid support and (ii) a capture reagent for a second analyte, wherein the capture reagent comprises a first binding member for a second analyte immobilized on a second different solid support, to produce a mixture. The mixture then is incubated for a time and under conditions sufficient to form capture reagent/first analyte and capture reagent/second

analyte complexes. The complexes so-formed then are contacted with a first indicator reagent comprising a member of a binding pair specific for the first analyte labeled with a detectable moiety and a second indicator reagent comprising a member of a binding pair specific for the second analyte labeled with a detectable moiety, to produce a second mixture. The second mixture then is incubated for a time and under conditions sufficient to produce both capture reagent/first analyte/first indicator reagent and capture reagent/second analyte/second indicator reagent complexes. The presence of one or more analytes is determined by detecting a signal generated by the complexes formed on either or both solid phases as an indication of the presence of one or more analytes in the test sample.

Other assay systems may employ an antibody which specifically binds US-type or US-subtype hepatitis E viral particles or sub-viral particles encapsulating the viral genome (or fragments thereof) by virtue of a contact between the specific antibody and the viral protein (peptide, etc.). The captured particles then can be analyzed by methods such as LCR or PCR to determine whether the viral genome is present in the test sample. The advantage of utilizing such an antigen capture amplification method is that it can separate the viral genome from other molecules in the test specimen by use of a specific antibody. Such a method has been described in EP 0 672 176, published September 20, 1995.

In general, immunoassay design considerations include preparation of antibodies (*e.g.*, monoclonal or polyclonal antibodies or antigen binding fragments thereof) having sufficiently high binding specificity for the target protein to form a complex that can be distinguished reliably from products of nonspecific interactions. Typically, the higher the antibody binding specificity, the lower the concentration of target that can be detected.

Both the polypeptide and antibody reagents of the invention may be used to develop assays as described herein to detect either the presence of an antigen from or an antibody that binds to a US-type or US-subtype hepatitis E virus. In addition to their use in immunoassays, it is contemplated that the aforementioned polypeptides may be used either alone or in combination with adjuvants for use in the production of antibodies in laboratory animals, or similarly, used in combination with pharmaceutically acceptable carriers as vaccines for either the prophylactic or therapeutic immunization of individuals. Also, it is contemplated that, in

addition to their use in immunoassays, the antibodies of the invention may be used in combination with, for example, a pharmaceutically acceptable carrier for use in passive, therapeutic or prophylactic immunization of an individual. These latter uses are described in more detail in section (III) below. The antibodies of the invention can also be used for the generation of chimeric antibodies for therapeutic use, or other similar applications.

Kits suitable for immunodiagnosis and containing the appropriate reagents may be constructed by packaging the appropriate materials, including, for example, a polypeptide defining a specific epitope of interest or antibodies that bind such epitopes in suitable containers. In addition, the kit optionally may include additional reagents, for example, suitable detection systems and buffers.

In addition, these antibodies, preferably monoclonal, can be bound to matrices similar to CNBr-activated Sepharose and used for the affinity purification of US-type or US-subtype hepatitis E specific proteins from cell cultures, or biological tissues such as blood and liver such as to purify recombinant and native viral antigens and proteins.

II. (ii) Nucleic Acid-based Assays

It is contemplated that if the marker is a US-type or US-subtype specific nucleotide sequence, the binding partner preferably also is a nucleotide sequence or an analog thereof that hybridizes specifically to the marker sequence or to regions adjacent thereto. Based on the unique polynucleotide sequences disclosed herein, it is contemplated that a binding partner may be a nucleotide sequence complementary to a US-type or US-subtype specific nucleotide sequence, for example, a nucleotide sequence or analog thereof complementary to at least a portion of an ORF 1 sequence, an ORF 2 sequence, or an ORF 3 sequence of a US-type or US-subtype hepatitis E virus, which is unique when compared to the corresponding nucleotide sequences of the Burmese and Mexican families. Furthermore, it is contemplated that noncoding portions of the genome of US-type and US-subtype hepatitis E viruses which are unique relative to the genomes of the Burmese and Mexican families of hepatitis E also may provide useful markers in the practice of the invention. Such nucleotide sequences (either

primers or probes) are of a length which allow detection of US-type or US-subtype specific sequences by hybridization and/or amplification and may be prepared using routine, standard methods, including automated oligonucleotide synthesis methodologies, well known and thoroughly discussed in the art. A complement of any unique portion of the HEV US-1 genome will be satisfactory. Complete complementarity is desirable for use as probes, although it may be unnecessary as the length of the fragment is increased.

Similarly, it is contemplated that the binding partner may be a polynucleotide sequence, for example, a DNA, RNA or PNA sequence, preferably comprising 8-100 nucleotides more preferably comprising 10-75 nucleotides and most preferably comprising 15-50 nucleotides, which is capable of hybridizing specifically to the target sequence. It is understood that the target sequence may be a nucleotide sequence defining at least a portion of a genome of a US-type or US-subtype hepatitis E virus, or a sequence complementary thereto. It is known in the art that the particular stringency conditions selected for a hybridization reaction depend largely upon the degree of complementarity of the binding partner nucleic acid sequence with the target sequence, the composition of the binding sequence and the length of the binding sequence. The parameters for determining stringency conditions are well known to those of ordinary skill in the art or are deemed to be readily ascertained from standard textbooks (see for example, Maniatis *et al.*, Molecular Cloning: A Laboratory Manual, (Cold Spring Harbor Press, N.Y., 1989)).

The sequences provided herein may be used to produce probes which can be used in assays for the detection of nucleic acids in test samples. The probes may be designed from conserved nucleotide regions of the polynucleotides of interest or from non-conserved nucleotide regions of the polynucleotide of interest. The design of such probes for optimization in assays is within the skill of the routineer. Generally, nucleic acid probes are developed from non-conserved or unique regions when maximum specificity is desired, and nucleic acid probes are developed from conserved regions when assaying for nucleotide regions that are closely related to, for example, different members of a multigene family or in related species like mouse and man.

One preferred protocol provides a method of detecting the presence or absence of a US-type or US-subtype hepatitis E virus in a test sample. The method comprises the steps of (a) providing a probe comprising a polynucleotide sequence containing at least 15 contiguous nucleotides from a US-type or US-subtype isolate, wherein the sequence is not present in other members of the hepatitis E Burmese and Mexican families; (b) contacting the test sample and the probe under conditions that permit formation of a polynucleotide duplex between the probe and its complement, in the absence of substantial polynucleotide duplex formation between the probe and non US-type and non US-subtype hepatitis polynucleotide sequences present in the test sample; and (c) detecting the presence of any polynucleotide duplexes containing the probe.

Preferred nucleotide sequences may comprise nucleotide residue numbers 1 through 5097 of SEQ ID NO:89, or a naturally occurring sequence variant thereof. With regard to this sequence, the term "a naturally occurring sequence variant" includes any nucleic acid sequence that is at least 73.3%, preferably at least 75.3%, more preferably at least 78.3%, and most preferably at least 95% identical to residues 1 through 5097 of SEQ ID NO:89. Other preferred marker or binding partner sequences may comprise nucleotide residue numbers 5132 through 7114 of SEQ ID NO:89, or a naturally occurring sequence variant thereof. With regard to this sequence, the term "naturally occurring sequence variant" includes any nucleic acid sequence that is at least 87.4%, preferably at least 89.4%, more preferably at least 92.4%, and most preferably at least 95% identical to residues 5132 through 7114 of SEQ ID NO:89. Other preferred marker or binding partner sequences may comprise nucleotide residue numbers 5094 through 5462 of SEQ ID NO:89, or a naturally occurring sequence variant thereof. With regard to this sequence, the term "naturally occurring sequence variant" includes any nucleic acid sequence that is at least 88.3% identical, preferably at least 90.3% identical, more preferably at least 93.3% identical, and most preferably at least 95% identical to residues 5094 through 5462 of SEQ ID NO:89.

Furthermore, it is contemplated that useful nucleotide sequences may include, for example, portions of the ORF 1 sequence encoding, for example, a protein selected from the group consisting of the methyltransferase protein, the protease protein, the Y domain protein, the X domain protein, the helicase protein, the hypervariable region protein and the RNA-

dependent RNA polymerase protein, or a variant thereof. Accordingly, it is contemplated that a useful methyltransferase encoding region of ORF 1 preferably has at least 78%, more preferably has at least 80%, and most preferably has at least 83% identity to residues 1-693 of SEQ ID NO:89. Also, it is contemplated that a useful protease encoding region of ORF 1

5 preferably has at least 66.1%, more preferably has at least 68.1%, and most preferably has at least 71.1% identity to residues 1270-2091 of SEQ ID NO:89. Also, it is contemplated that a useful Y domain encoding region of ORF 1 has at least 80%, more preferably has at least 82%, and most preferably has at least 85% identity to residues 619-1272 of SEQ ID NO:89. Also, it is contemplated that a useful X domain encoding region of ORF 1 has at least 73.5%, more

10 preferably has at least 75.5%, and most preferably has at least 78.5% identity to residues 2365-2841 of SEQ ID NO:89. Also, it is contemplated that a useful helicase encoding region of ORF 1 has at least 77.5%, and most preferably has at least 79.5%, and most preferably has at least 81.5% identity to residues 2893-3591 of SEQ ID NO:89. Also, it is contemplated that a useful hypervariable region encoding region of ORF 1 has at least 51.2%, more preferably has at least

15 53.2%, and most preferably has at least 56.2% identity to residues 2092-2364 of SEQ ID NO:89. Also, it is contemplated that a useful RNA-dependent RNA polymerase encoding region of ORF 1 has at least 76.3%, more preferably has at least 78.3%, and most preferably has at least 81.3% identity to residues 3634-5094 of SEQ ID NO:89.

Preferred nucleotide sequences may comprise nucleotide residue numbers 36 through

20 5162 of SEQ ID NO:164, or a naturally occurring sequence variant thereof. With regard to this sequence, the term "a naturally occurring sequence variant" includes any nucleic acid sequence that is at least 73.6%, preferably at least 75.6%, more preferably at least 78.6% and more preferably at least 95% identical to residues 36 through 5162 of SEQ ID NO:164. Other preferred marker or binding partner sequences may comprise nucleotide residue numbers 5197

25 through 7179 of SEQ ID NO:164, or a naturally occurring sequence variant thereof. With regard to this sequence, the term "naturally occurring sequence variant" includes any nucleic acid sequence that is at least 80.7%, preferably at least 82.7%, more preferably at least 85.7% and most preferably at least 95% identical to residues 5197 through 7179 of SEQ ID NO:164. Other preferred marker or binding partner sequences may comprise nucleotide residue numbers

30 5159 through 5527 of SEQ ID NO:164, or a naturally occurring sequence variant thereof. With

regard to this sequence, the term “naturally occurring sequence variant” includes any nucleic acid sequence that is at least 87.9% identical, preferably at least 89.9% identical, more preferably at least 92.9% identical and even more preferably at least 95% identical to residues 5159 through 5527 of SEQ ID NO:164.

Furthermore, it is contemplated that useful HEV US-2 nucleotide sequences may include, for example, portions of the ORF 1 sequence encoding, for example, at least a portion of a protein selected from the group consisting of the methyltransferase protein, the protease protein, the Y domain protein, the X domain protein, the helicase protein, the hypervariable region protein and the RNA-dependent RNA polymerase protein, or a variant thereof.

Accordingly, it is contemplated that a useful methyltransferase encoding region of ORF 1 preferably has at least 79.5%, more preferably has at least 81.5%, and most preferably has at least 84.5% identity to residues 36-755 of SEQ ID NO:164. Also, it is contemplated that a useful protease encoding region of ORF 1 preferably has at least 66.1%, more preferably has at least 68.1%, and most preferably has at least 71.1% identity to residues 1332-2153 of SEQ ID NO:164. Also, it is contemplated that a useful Y domain encoding region of ORF 1 has at least 80.7%, more preferably has at least 82.7%, and most preferably has at least 85.7% identity to residues 680-1334 of SEQ ID NO:164. Also, it is contemplated that a useful X domain encoding region of ORF 1 has at least 73.7%, more preferably has at least 75.7%, and most preferably has at least 78.7% identity to residues 2430-2906 of SEQ ID NO:164. Also, it is contemplated that a useful helicase encoding region of ORF 1 has at least 76.4%, and most preferably has at least 78.4%, and most preferably has at least 81.4% identity to residues 2958-3656 of SEQ ID NO:164. Also, it is contemplated that a useful hypervariable region encoding region of ORF 1 has at least 50.4%, more preferably has at least 52.8%, and most preferably has at least 55.8% identity to residues 2154-2429 of SEQ ID NO:164. Also, it is contemplated that a useful RNA-dependent RNA polymerase encoding region of ORF 1 has at least 76.8%, more preferably has at least 78.8%, and most preferably has at least 81.8% identity to residues 3699-5159 of SEQ ID NO:164.

Other useful nucleotide sequences comprise the nucleotide sequences that encode the amino acid sequences selected from the group consisting of SEQ ID NOS:93, 168, 173, 174, 175, 176, 223, and 224 and nucleotide sequences complementary thereto.

It is contemplated that the nucleic acid sequences provided herein may be used to determine the presence of US-type or US-subtype hepatitis E virus in a test sample by conventional nucleic acid based assays, for example, by polymerase chain reaction (PCR) and/or by blot hybridization studies (described in detail below). In addition to their use in nucleic acid based assays, it is contemplated the aforementioned nucleic acid sequences may be integrated in vectors which may then be transformed or transfected into a host cell of interest, for example, vaccinia or mycobacteria. The resulting host cells may then be combined with a pharmaceutically acceptable carrier and used, for example, as a recombinant vaccine for immunizing a mammal, either prophylactically or therapeutically, against a preselected US-type or US-subtype hepatitis E virus.

The polymerase chain reaction (PCR) is a technique for amplifying a desired nucleic acid sequence (target) contained in a nucleic acid or mixture thereof. In PCR, a pair of primers typically are employed in excess to hybridize at the outside ends of complementary strands of the target nucleic acid. The primers are each extended by a polymerase, for example, a thermostable polymerase, using the target nucleic acid as a template. The extension products become target sequences themselves, following dissociation from the original target strand. New primers then are hybridized and extended by a polymerase, and the cycle is repeated to geometrically increase the number of target sequence molecules. PCR is disclosed in U.S. patents 4,683,195 and 4,683,202.

The Ligase Chain Reaction (LCR) is an alternate method for nucleic acid amplification. In LCR, probe pairs are used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess of the target nucleic acid sequence. The first probe hybridizes to a first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In

addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Once the ligated strand of primary probes is separated from the target strand, it will hybridize with the third and fourth probes which can be ligated to form a complementary, secondary
 5 ligated product. The ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. This technique is described more completely in EP-A- 320 308 to K. Backman published June 16, 1989 and EP-A-439 182 to K. Backman *et al.*, published July 31, 1991.

10 For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No. 5,322,770; or to reverse transcribe mRNA into cDNA followed by asymmetric gap ligase chain reaction (RT-AGLCR) as described by R. L. Marshall, *et al.*, PCR Methods and Applications 4: 80-84 (1994).

15 Other known amplification methods which can be utilized herein include but are not limited to the so-called "NASBA" or "3SR" technique described in Proc. Natl. Acad. Sci. USA 87: 1874-1878 (1990) and also described in Nature 350 (No. 6313): 91-92 (1991); Q-beta amplification as described in published EP 4544610; strand displacement amplification (as described in G. T. Walker *et al.*, Clin. Chem. 42: 9-13 [1996]) and EP 684315; and target
 20 mediated amplification, as described by PCT Publication WO 9322461.

In one embodiment, the present invention generally comprises the steps of contacting a test sample suspected of containing a target polynucleotide sequence with amplification reaction reagents comprising an amplification primer, and a detection probe that can hybridize with an internal region of the amplicon sequences. Probes and primers employed according to
 25 the method herein provided are labeled with capture and detection labels wherein probes are labeled with one type of label and primers are labeled with the other type of label. Additionally, the primers and probes are selected such that the probe sequence has a lower melt temperature than the primer sequences. The amplification reagents, detection reagents and test sample are placed under amplification conditions whereby, in the presence of target sequence,

copies of the target sequence (an amplicon) are produced. The double stranded amplicon then is thermally denatured to produce single stranded amplicon members. Upon formation of the single stranded amplicon members, the mixture is cooled to allow the formation of complexes between the probes and single stranded amplicon members.

5 After the probe/single stranded amplicon member hybrids are formed, they are detected. Standard heterogeneous assay formats are suitable for detecting the hybrids using the detection labels and capture labels present on the primers and probes. The hybrids can be bound to a solid phase reagent by virtue of the capture label and detected by virtue of the detection label. In cases where the detection label is directly detectable, the presence of the hybrids on the solid
10 phase can be detected by causing the label to produce a detectable signal, if necessary, and detecting the signal. In cases where the label is not directly detectable, the captured hybrids can be contacted with a conjugate, which generally comprises a binding member attached to a directly detectable label. The conjugate becomes bound to the complexes and the conjugates presence on the complexes can be detected with the directly detectable label. Thus, the
15 presence of the hybrids on the solid phase reagent can be determined. Those skilled in the art will recognize that wash steps may be employed to wash away unhybridized amplicon or probe as well as unbound conjugate.

Test samples for detecting target sequences can be prepared using methodologies well known in the art such as by obtaining a sample and, if necessary, disrupting any cells contained
20 therein to release target nucleic acids. In the case where PCR is employed in this method, the ends of the target sequences are usually known. In cases where LCR or a modification thereof is employed in the preferred method, the entire target sequence is usually known. Typically, the target sequence is a nucleic acid sequence such as, for example, RNA or DNA.

While the length of the primers and probes can vary, the probe sequences are selected
25 such that they have a lower melt temperature than the primer sequences. Hence, the primer sequences are generally longer than the probe sequences. Typically, the primer sequences are in the range of between 20 and 50 nucleotides long, more typically in the range of between 20 and 30 nucleotides long. Preferred primer sequences typically are greater than 20 nucleotides long. The typical probe is in the range of between 10 and 25 nucleotides long more typically in

the range of between 15 and 20 nucleotides long. Preferred probe sequences typically are greater than 15 nucleotides long.

Alternatively, a probe may be involved in the amplifying a target sequence, via a process known as "nested PCR". In nested PCR, the probe has characteristics which are similar to those of the first and second primers normally used for amplification (such as length, melting temperature etc.) and as such, may itself serve as a primer in an amplification reaction.

Generally in nested PCR, a first pair of primers (P₁ and P₂) are employed to form primary extension products. One of the primary primers (for example, P₁) may optionally be a capture primer (i.e. linked to a member of a first reactive pair), whereas the other primary primer (P₂) is not. A secondary extension product is then formed using a probe (P₁') and a probe (P₂') which may also have a capture type label (such as a member of a second reactive pair) or a detection label at its 5' end. The probes are complementary to and hybridize at a site on the template near or adjacent the site where the 3' termini of P₁ and P₂ would hybridize if still in solution.

Alternatively, a secondary extension product can be formed using the P₁ primer with the probe (P₂') or the P₂ primer with the probe (P₁') sometimes referred to as "hemi-nested PCR". Thus, a labeled primer/probe set generates a secondary product which is shorter than the primary extension product. Furthermore, the secondary product may be detected either on the basis of its size or via its labeled ends (by detection methodologies well known to those of ordinary skill in the art). In this process, probe and primers are generally employed in equivalent concentrations.

Various methods for synthesizing primers and probes are well known in the art. Similarly, methods for attaching labels to primers or probes are also well known in the art. For example, it is a matter of routine experimentation to synthesize desired nucleic acid primers or probes using conventional nucleotide phosphoramidite chemistry and instruments available from Applied Biosystems, Inc., (Foster City, CA), Dupont (Wilmington, DE), or Milligen (Bedford MA). Many methods have been described for labeling oligonucleotides such as the primers or probes of the present invention. Enzo Biochemical (New York, NY) and Clontech (Palo Alto, CA) both have described and commercialized probe labeling techniques. For

example, a primary amine can be attached to a 3' oligo terminus using 3'-Amine-ON CPGTM (Clontech, Palo Alto, CA). Similarly, a primary amine can be attached to a 5' oligo terminus using Aminomodifier II[®] (Clontech). The amines can be reacted to various haptens using conventional activation and linking chemistries. In addition, WO 92/10506, published 25 June 5 1992 and U. S. Patent 5,290,925, issued March 1, 1994, teach methods for labeling probes at their 5' and 3' termini, respectively. In addition, WO 92/11388 published 9 July 1992 teaches methods for labeling probes at their ends. According to one known method for labeling an oligonucleotide, a label-phosphoramidite reagent is prepared and used to add the label to the oligonucleotide during its synthesis. See, for example, N.T. Thuong *et al.*, Tet. Letters 29(46): 10 5905-5908 (1988); or J. S. Cohen *et al.*, published U.S. Patent Application 07/246,688 (NTIS ORDER No. PAT-APPL-7-246,688) (1989). Preferably, probes are labeled at their 3' and 5' ends.

Capture labels are carried by the primers or probes and can be a specific binding member which forms a binding pair with the solid phase reagent's specific binding member. It will be understood, of course that the primer or probe itself may serve as the capture label. For 15 example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of the primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where the probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or 20 "tail" that is not complementary to the single stranded amplicon members. In the case where the primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase because the probe is selected such that it is not fully complementary to the primer sequence.

Generally, probe/single stranded amplicon member complexes can be detected using 25 techniques commonly employed to perform heterogeneous immunoassays. Preferably, in this embodiment, detection is performed according to the protocols used by the commercially available Abbott LCx[®] instrumentation (Abbott Laboratories, Abbott Park, IL).

Other useful procedures known in the art include solution hybridization, and dot and slot blot hybridization protocols. The amount of the target nucleic acid present in a sample optionally may be quantitated by measuring the radioactivity of hybridized fragments, using standard procedures known in the art.

5 **III. Vaccines**

It is contemplated that vaccines may be prepared from one or more immunogenic polypeptides based on US-type and/or US-subtype specific protein sequences or antibodies that bind to such protein sequences. In addition, it is contemplated that vaccines also may comprise dead, live but attenuated US-type or US-subtype hepatitis E virus, or a live, recombinant
10 vaccine comprising a heterologous host cell, for example, a vaccinia virus, expressing a US-type or US-subtype hepatitis E virus specific antigen.

With regard to the polypeptide based vaccines, the polypeptide must define at least one epitope. It is contemplated, however, that the vaccine may comprise a plurality of different epitopes which are defined by one or more polypeptide chains. Furthermore, it is contemplated
15 that nonstructural proteins as well as structural proteins may provide protection against viral pathogenicity, even if they do not cause the production of neutralizing antibodies. Considering the above, multivalent vaccines against the US-type or US-subtype virus may comprise one or more structural proteins, and/or one or more nonstructural proteins. These immunogenic epitopes can be used in combinations, *i.e.*, as a mixture of recombinant proteins, synthetic
20 peptides and/or polypeptides isolated from the virion; which may be co-administered at the same or administered at different time.

Methodologies for the preparation of protein or peptide based vaccines which contain at least one immunogenic peptide as an active ingredient are well known in the art. Typically, such vaccines are prepared as injectables, either as liquid solutions or suspensions. The
25 preparation may be emulsified or the protein may be encapsulated in liposomes. The active immunogenic ingredients may be mixed with pharmacologically acceptable excipients which are compatible with the active ingredient. Suitable excipients include, without limitation, water, saline, dextrose, glycerol, ethanol or a combination thereof. The vaccine also may

contain small amounts of auxiliary substances such as wetting or emulsifying reagents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. For example, such adjuvants can include aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-DMP), N-acetyl-nomuramyl-L-alanyl-D-isoglutamine (CGP 11687, also referred to as nor-MDP), N-acetyl-muramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'2'-dipalmitoyl sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, also referred to as MTP-PE), and RIBI (MPL + TDM + CWS) in a 2% squalene/Tween-80® emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic polypeptide containing a US-type or US-subtype specific antigenic sequence resulting from administration of this polypeptide in vaccines which also comprise various adjuvants under investigation.

The vaccines usually are administered by intravenous or intramuscular injection. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include but are not limited to polyalkylene glycols or triglycerides. Such suppositories may be formed from mixtures containing the active ingredient in the range of from about 0.5% to about 10%, preferably, from about 1% to about 2% (w/w). Oral formulation may include excipients including, for example, mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate and the like. These compositions may take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain about 10% to about 95% of active ingredient, preferably about 25% to about 70% (w/w).

The polypeptide chains used in the vaccine may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include, for example, acid addition salts formed by the addition of inorganic acids such as hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, or other acids known to those skilled in the art. Salts formed with the free carboxyl groups also may be derived from inorganic bases such as sodium, potassium, ammonium, calcium or ferric hydroxides and the like, and organic

bases such as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine procaine, or other bases known to those skilled in the art.

Vaccines typically are administered in a way compatible with the dosage formulation, and in such amounts that will be effective prophylactically and/or therapeutically. The quantity to be administered generally ranges from about 5 μ g to about 250 μ g of antigen per dose, however the actual dose will depend upon the health and size of the subject, the capacity of the subject's immune system to synthesize antibodies, and the degree of protection sought. The vaccine may be given in a single or multiple dose schedule. A multiple dose is one in which a primary course of vaccination may be with one to ten separate doses, followed by other doses given at subsequent time intervals required to maintain and/or to reinforce the immune response, for example, at one to four months for a second dose, and if required by the individual, a subsequent dose(s) several months later. In addition, the dosage regimen may be determined, at least in part, by the need of the individual, and may be dependent upon the practitioner's judgment.

With regard to dead or otherwise inactivated US-type or US-subtype hepatitis E virus containing vaccines, inactivation may be facilitated using conventional methodologies well known and thoroughly documented in the art. Preferred inactivation methods include, for example, exposure to one or more of (i) organic solvents, (ii) detergents, (iii) formalin, and (iv) ionizing radiation. It is contemplated that some of the proteins in attenuated vaccines may cross-react with other known viruses, and thus shared epitopes may exist between a US-type or US-subtype hepatitis E virus and other members of the HEV family (for example, members of the Burmese or Mexican families) and thus give rise to protective antibodies against one or more of the disorders caused by these pathogenic agents. Preferred formulations and modes of administration are thoroughly documented in the art and so are not discussed in detail herein. The various factors to be considered may include one or more features discussed hereinabove for the peptide based vaccines.

With regard to the live, but attenuated vaccines, it may be possible to produce attenuated virus using any of the attenuation methods known and used in the art. Briefly, attenuation may be accomplished by passage of the virus at low temperatures or by introducing

missense mutations or deletions into the viral genome. Preferred formulations and modes of administration are thoroughly documented in the art and so are not discussed in detail herein. The various factors to be considered may include one or more features discussed hereinabove for the peptide based vaccines.

5 With regard to live, recombinant vaccines (vector vaccines), these may be developed by incorporating into the genome of a living but harmless virus or bacterium, a gene or nucleic acid sequence encoding a US-type or US-subtype hepatitis E specific polypeptide chain defining an antigenic determinant. The resulting vector organism may then be administered to the intended host. Typically, for such a vaccine to be successful, the vector organism must be viable, and
10 either naturally non-virulent or have an attenuated phenotype. Preferred host organisms include, vaccinia virus, adenovirus, adeno-associated virus, salmonella and mycobacteria. Live strains of vaccinia virus and mycobacteria have been administered safely to humans in the forms of the smallpox and tuberculosis (BCG) vaccines, respectively. In addition, they have been shown to express foreign proteins and exhibit little or no conversion into virulent phenotypes. Vector
15 vaccines are capable of carrying a plurality of foreign genes or nucleic acid sequences thereby permitting simultaneous vaccination against a variety of preselected antigenic determinants. Preferred formulations and modes of administration are thoroughly documented in the art and so are not discussed in detail herein.

20 **IV. Identification of molecules with anti-US-type or anti-US-subtype hepatitis E virus activity.**

In view of the discovery of specific HEV US-type sequences, it is contemplated that one skilled in the art may be able to identify molecules which either inactivate or reduce the activity of HEV US-type specific proteins, *e.g.*, the helicase, methyltransferase, or protease proteins
25 encoded by the ORF 1 portions of the HEV genome. An exemplary protocol for identifying molecules that inhibit the HCV protease is described in U.S. Patent No. 5,597,691, the disclosure of which is incorporated herein by reference. Although, the method pertains to the

identification of HCV protease inhibitors, it is contemplated that the same or similar protocols maybe used to identify HEV protease inhibitors, or any other protein encoded by a HEV US-type sequence.

Briefly, a method for identifying HEV protease inhibitors is as follows. Typically, a
5 substrate is employed which mimics the proteases natural substrate, but which provides a quantifiable signal when cleaved. The signal preferably is detectable by colorimetric or fluorometric means; however, other methods such as HPLC or silica gel chromatography, nuclear magnetic resonance, and the like may also be useful. After optimum substrate and protease concentrations have been determined, candidate protease inhibitors are added one at a
10 time to the reaction mixture at a range of concentrations. The assay conditions preferably resemble the conditions under which the protease is to be inhibited *in vivo*, *i.e.*, under physiologic pH, temperature, ionic strength, etc. Suitable inhibitors exhibit strong protease inhibition at concentrations which do not raise toxic side effects in the subject. Inhibitors which compete for binding to the protease active site may require concentrations equal to or
15 greater than the substrate concentration, while inhibitors capable of binding irreversibly to the protease active site may be added in concentrations on the order of the enzyme concentration.

It is contemplated that the inhibitors may be organic compounds, which, for example, mimic the cleavage site recognized by the HEV protease, or alternatively, may be proteins, for example, antibodies or antibody fragments capable of binding specifically to and inactivating or
20 reducing the activity of the HEV protease. Once identified, the protease inhibitors may be administered by a variety of methods, such as intravenously, orally, intramuscularly, intraperitoneally, bronchially, intranasally, and so forth. The preferred route of administration will depend upon the nature of inhibitor. Inhibitors prepared as organic compounds may be administered orally (which is generally preferred) if well absorbed. Protein-based inhibitors
25 (such as most antibodies or antibody derivatives) generally are administered by parenteral routes.

Examples

Practice of the invention will be more fully understood from the following examples, which are presented herein for illustrative purposes only, and should not be construed as limiting the invention in any way. All citations to the literature, both *supra* and *infra*, including patents, patent applications and scientific publications are incorporated by reference herein, in their entirety.

Example 1 – Case study

HEV strain US-1 was identified in the serum of a patient (USP-1) suffering from acute hepatitis. The patient was a 62 year old, white male who was hospitalized in Rochester, MN after a three-week history of fever, abdominal pain, jaundice, and pruritis. Onset of signs and symptoms began two weeks after returning home following a ten day trip to San Jose, California.

His past medical history included a nephrectomy for autosomal dominant polycystic kidney disease accompanied by mild renal insufficiency, and a laparoscopic cholecystectomy for symptomatic cholelithiasis. The patient had osteoarthritis and was hypertensive. Lisinopril therapy had been initiated three months prior to admission. Physical examination revealed an ill appearing icteric white male with an enlarged tender liver, and no asterixis. Serum aspartate aminotransferase (AST), alanine aminotransferase (ALT), and bilirubin levels were markedly elevated at the time of hospital admission and peaked 8 days and 16 days after hospitalization, respectively (Figure 2). Lisinopril was discontinued on admission. Serologies for hepatitis A (IgM and IgG anti-HAV), hepatitis B (HBsAg, IgM and IgG anti-HBc), hepatitis C (anti-HCV), and HCV RNA were negative. Ceruloplasmin, iron, transferrin, anti-nuclear and anti-smooth muscle antibodies, toxin and drug screen were all normal. Careful questioning of the patient revealed no history of ethanol use. Abdominal ultrasound and computed tomography scan, and endoscopic retrograde cholangiopancreatogram were also normal. A liver biopsy showed a severe, acute lobular hepatitis with striking pyknotic and ballooning degeneration of

hepatocytes consistent with autoimmune, drug, or viral hepatitis.

The patient made a complete clinical recovery within 2 months, with normalization of AST, ALT, and bilirubin noted about 5 months after hospital admission. No risk factors for acquiring HEV were identified. He had not traveled outside the US for over 10 years. In the 6 weeks prior to illness onset, the only meals he reported eating that were not prepared at home were at a Mexican restaurant and a large fast food restaurant chain. He had no exposure to untreated drinking water, did not report eating raw shellfish, and had no known exposure to farm animals. None of the food handlers at the Mexican restaurant or the fast food restaurant reported foreign travel since less than 5 months from admission date and none reported signs and/or symptoms of hepatitis. No other cases of non-ABC hepatitis were reported in the county health department where the patient stayed in California, and where the patient lived in Minnesota during the period of admission. No family members had signs and/or symptoms of hepatitis either during the patient's trip to California or in the subsequent 10 weeks. Serum obtained from 6 family members in California, and from his spouse who lived with him in Minnesota over the period of interest were negative for anti-HEV by EIA.

Example 2 – Identification of unique isolate of HEV US-1

The presence of HEV was determined by RT-PCR using HEV primer sequences. Briefly, nucleic acids were isolated from 25 μ L of serum from patient USP-1 as previously described (Schlauder *et al.* (1995) J. Virological Methods 46: 81-89). Ethanol precipitated nucleic acids were resuspended in 3 μ L of diethyl pyrocarbonate (DEPC) treated water.

cDNA synthesis and PCR were performed using the GeneAmp RNA PCR kit from Perkin-Elmer (Norwalk, CT) in accordance with the manufacturer's instructions. RNA (1 μ L) was used as a template for each 10 μ L cDNA reaction. cDNA synthesis was primed with specific primers added to a final concentration of 4 μ M. The subsequent amplification of cDNA was primed with oligonucleotides added to a final concentration of 0.8 to 1.0 μ M. PCR was performed for 40 cycles (94°C, 20 sec; 55°C, 30 sec; 72°C, 30 sec; followed by an extension cycle of 72°C for 3 min). The initial PCR reaction (2 μ L) then was used as a template for a second round of amplification using a nested set of PCR primers. PCR was performed using

the GeneAmp PCR kit from Perkin-Elmer in accordance with the manufacturer's instructions. Briefly, primers were added to a final concentration of 1 μ M. The initial set of experiments used three sets of primers. Two from the 5'-end of ORF 1 based on sequences from the Burmese and Mexican strains. One set from the 3'-end of ORF 1 based on the Mexican strain sequence. The three sets of primers used were as follows:

Primer Set 1

<u>Primer</u>	<u>Sequence</u>	<u>SEQ ID NO:</u>
5'-ORF 1-Mexican primer C375M	CTGAACATCCCGGCCGAC	SEQ ID NO:1
PCR primer A1-350M	AGAAAGCAGCGATGGAGGA	SEQ ID NO:2
PCR primer S1-34M	GCCCACCAGTTCATTAAGGCT	SEQ ID NO:3
nested PCR primer A2-320M	TCATTAATGGAGCGTGGGTG	SEQ ID NO:4
nested PCR primer S2-55M	CCTGGCATCACTACTGCTAT	SEQ ID NO:5

Primer Set 2

<u>Primer</u>	<u>Sequence</u>	<u>SEQ ID NO:</u>
5'-ORF 1- Burmese cDNA primer C375	CTGAACATCACGCCCAAC	SEQ ID NO:6
PCR primer A1-350	AGGAAGCAGCGGTGGACCA	SEQ ID NO:7
PCR primer S1-34	GCCCATCAGTTTATTAAGGC	SEQ ID NO:8
nested PCR primer A2-320	TCATTTATTGAGCGGGGATG	SEQ ID NO:9
nested PCR primer S2-55	CCTGGCATCACTACTGCTAT	SEQ ID NO:10

Primer Set 3

<u>Primer</u>	<u>Sequence</u>	<u>SEQ ID NO:</u>
3'-ORF 1- Mexican cDNA primer M1PR6	CCATGTTCCACACCGTATTCCAGAG	SEQ ID NO:11
PCR primer S4294M	GTGTTCTACGGGGATGCTTATGACG	SEQ ID NO:12
nested PCR primer M1PF6	GACTCAGTATTCTCTGCTGCCGTGG	SEQ ID NO:13
nested PCR primer A4556	GGCTCACCAGAATGCTTCTTCCAGA	SEQ ID NO:14

The 5'-ORF 1-Burmese primers are described in Schlauder *et al.* (1993) *Lancet* 341: 378. Primers M1PR6 and M1PF6 are described in McCaustland *et al.* (1991) *J. Virological Methods* 35: 331-342. The PCR products were separated by agarose gel electrophoresis and visualized by UV irradiation after ethidium bromide staining. The resulting PCR products were
 5 hybridized to a radiolabelled probe after Southern blot transfer to a nitrocellulose filter.

Radiolabelled probes were generated from PCR products purified with the QIAEX gel extraction purification kit by Qiagen (Chatsworth, CA). Radiolabel was incorporated using the Stratgene® (La Jolla, CA) Prime-It II kit according to the manufacturer's instructions. Filters were prehybridized in Rapid-hyb buffer from Amersham (Arlington Heights, IL) for 3-5 hours,
 10 and then hybridized in Fast-Pair Hybridization Solution with 100-200 cpm/cm² at 42°C for 15-25 hours. Filters then were washed as described in Schlauder *et al.* (1992) *J. Virol. Methods* 37: 189-200. Phosphorimages of the probed filters were obtained with a Molecular Dynamics Phosphorimager 425E (Sunnyvale, CA).

Ethidium bromide stained bands were detected with the primers from the 5'-end of ORF
 15 1. However, only the primers based on the Mexican strain resulted in a nested product of the expected size of 266 base pairs. Hybridization to a probe derived from a Burmese-like strain (identity > 90%) infected patient resulted in a very weak hybridization signal to the patient USP-1 derived products relative to the signal from the Burmese positive control. These results gave the first indication that this isolate was not closely related to the Burmese isolate. No
 20 probe was available from the Mexican strain.

To confirm these results, RNA was extracted from additional serum aliquots of patient USP-1. RT-PCR was performed using the 5'-ORF 1-Mexican primers, SEQ ID NOS:1-5, as described above. Following agarose gel electrophoresis and staining with ethidium bromide, a 342 bp product was visualized in each sample. The PCR products were extracted from the
 25 agarose gel using the QIAEXII Agarose Gel Extraction Kit by Qiagen (Chatsworth, CA) and cloned into pT7 Blue T-vector plasmid by Novagen (Madison, WI). The cloned products were sequenced using the SEQUENASE VERSION 2.0 sequencing kit (USB, Cleveland, OH) in accordance with the manufacturers instructions.

The nucleotide sequences obtained from the product of the latter two samples were identical and are shown in SEQ ID NO:15. These results indicate that only the cDNA primer and primer S1 from both the Burmese and Mexican strains resulted in an ethidium bromide stainable product from the patient USP-1 samples. Only the Mexican strain based nested primers, S2 and A2 generated an ethidium bromide stainable product of the expected size.

In order to determine the degree of relatedness between the HEV US-1 isolate and other known isolates of HEV, alignments of the nucleotide and amino acid sequences were performed using the program GAP of the Wisconsin Sequence Analysis Package (Version 9), available from the Genetics Computer Group, Inc., 575 Science Drive, Madison, Wisconsin, 53711. The program employs the algorithm of Needleman and Wunsch (J. Mol. Biol. (1970) 48:443-453) to calculate the degree of similarity and identity, which are expressed as percentages between the two sequences being aligned. The gap creation and gap extension penalties were 50 and 3.0, respectively, for nucleic acid sequence alignments, and 12 and 4, respectively, for amino acid sequence comparisons.

The complete nucleotide and amino acid sequences of the two 'prototype' HEV isolates from Burma and Mexico, as well as other sequences used for analyses were obtained from GenBank, with their respective accession numbers are indicated in Table 1 below. Each of the these sequences are incorporated herein by reference.

TABLE 1

Isolate	Genbank Accession Number
Mexican (M1)	M74506
Burmese (B1)	M73218
Pakistan (P1)	M80581
Chinese (C4)	D11093

A 303 base pair sequence of HEV US-1 (homologous to residues 1-303 of SEQ ID NO:89) was compared against the homologous regions identified in the Mexican, Burmese,

Pakistani, and Chinese strains. The resulting percent identities are summarized in Table 2 below.

TABLE 2. Identity over 303 nucleic acids from the 5'-end ORF 1 product

	US-1	Mexican	Burmese	Pakistan
Mexican	77.2			
Burmese	74.9	83.2		
Pakistan	75.9	83.2	95.7	
Chinese	75.9	83.5	95.7	97.4

The results in Table 2 indicate that the fragment from the 5'-end of ORF 1 from the USP-1 isolate showed a nucleic acid identity from about 74.9 to about 77.2 % relative to other known isolates of HEV. This was less than the identity between the prototype Mexican and Burmese isolates (83.2%). These results indicate that the product likely was derived from a unique isolate of HEV not previously identified.

Example 3 – Genome Extension and Sequencing of HEV US-1

The clone obtained and sequenced as described in Example 2 (SEQ ID NO:15) hereinabove was derived from a unique HEV genome, HEV US-1. To obtain sequences from additional regions of the HEV US-1 genome, several reverse transcriptase-polymerase chain reaction (RT-PCR) walking experiments were performed.

Total nucleic acids were extracted by the procedure described in Example 2 (for SEQ ID NO:19 only) or by one of the following procedures. Aliquots (25 μ L) of patient USP-1 serum were extracted using the Total Nucleic Acid Extraction procedure in accordance with the manufacturers instructions (United States Biochemical) in the presence of 10 mg yeast tRNA as carrier. Nucleic acids were precipitated and resuspended in 3.75 μ L RNase/DNase free water. Alternatively, total RNA was isolated from 100 μ L of serum using the ToTALLY RNA isolation kit as recommended by the manufacturer (Ambion, Inc.). The resulting RNAs were treated with DNase and column purified with reagents from S.N.A.P. Total RNA isolation kit (Invitrogen, San Diego, CA). Thereafter, RNA was precipitated with 0.1 volumes of 3M

sodium acetate, 2 μ L pellet paint (Novagen) as carrier, and 2 volumes ethanol. RNA pellets were dissolved in 50 μ L DEPC treated water.

RT-PCR was performed using the-GeneAmp RNA PCR kit in accordance with the manufacturers instructions (Perkin-Elmer). Random hexamers were used to prime cDNA synthesis in a total volume of 25 μ L except for the isolation of SEQ ID NO:19 which utilized cDNA specifically primed with primer PA2-5560 (SEQ ID NO:16), as described in Example 2 above. US1-gap was generated with specifically primed cDNA generated using RNA extracted from 12.5 μ L serum equivalents, primer US1 gap-a0.5 (SEQ ID NO:46), and Superscript II (3' RACE Kit: GIBCO BRL). PCR was performed with the cDNA encompassing one-fifth of the total reaction volume (2 μ L for 10 μ L reaction or 5 μ L for 25 μ L reaction, etc.). Standard PCR was performed in the presence of 2 mM $MgCl_2$ and 0.5 to 1.0 μ M of each primer. Modified reactions contained 1x PCR Buffer and 20% Q Solution (Qiagen) in accordance with the manufacturer's instructions for the isolation of SEQ ID NOS:33 and 41. Reactions used two HEV consensus primers (Table 3), one HEV consensus primer and one HEV-US-1 specific primer (Table 4), two HEV US-1 specific primers (Table 5), one HEV US-1 specific primer and one HEV US-2 (see Example 5) specific primer (Table 6), or two HEV US-2 specific primers (Table 7). Reactions were subjected to thermal cycling as follows:

SEQ ID NOS:19, 24, 27, 30, 33, 41, 44, 60, 64, 68, 73, 78, and 83 were obtained by touchdown PCR. Amplification involved 43 cycles of 94°C for 30 seconds, 55°C for 30 seconds (-0.3°C/cycle), and 72°C for 1 minute. This was followed by 10 cycles of 94°C for 30 seconds, 40°C for 30 seconds, and 72°C for 1 minute. For SEQ ID NOS:38, 49, 52, and 55, cycling involved 35 rounds of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1 minute. All amplifications were preceded by 1-2 minutes at 94°C and followed by 72°C for 5 to 10 minutes. The reactions were held at 4°C prior to agarose gel analysis.

The isolation of SEQ ID NO:19 required a second round of touch down amplification to isolate the desired product. Here, 1 μ L of first round was placed into a second round 25 μ L reaction. The second round amplification utilized hemi-nested primers as indicated in Table 3 by reactions 1.1.1 and 1.1.2. The isolation of SEQ ID NO:24 required a second round of nested

touch down amplification as described above and indicated in Table 4 as reactions 2.1.1 and 2.1.2. The isolation of SEQ ID NOS:38 and 49 required a second round of nested PCR (Table 5) utilizing 1 μ L of first round into a 25 μ L reaction as described above. The isolation of SEQ ID NOS:60, 64, 68, and 73 required nested PCR in which 1 μ L of the first round was amplified in a 25 μ L second round reaction (Table 6). Products SEQ ID NOS:78 and 83 were generated from two rounds of amplification (Table 7).

Agarose gel electrophoresis was performed on a fraction or all of the PCR reaction in a 0.8% to 2% agarose TAE gel in the presence of 0.2 mg/mL ethidium bromide. Products were visualized by UV irradiation and products of the desired molecular weight were excised, purified using GeneClean in accordance with the manufacturers' instructions (BIO 101, Inc.), and cloned into pT7-Blue T-Vector plasmid (Novagen) II or pGEM-T Easy Vector (Promega) in accordance with the manufacturers' instructions. Cloned products were sequenced as described in Example 2 or on a ABI Model 373 DNA Sequencer using ABI Sequencing Ready Reaction Kit as specified by the manufacturer. Results of these experiments are presented hereinbelow in Tables 3, 4, 5, 6, and 7.

TABLE 3

Reaction	Primer 1	Primer 2	Approx. Prod. Size/SEQ ID
1.1.1	SEQ ID NO:17	SEQ ID NO:16	none
1.1.2	SEQ ID NO:18	SEQ ID NO:16	251 bp/SEQ ID NO:19
1.2	SEQ ID NO:28	SEQ ID NO:29	168 bp/SEQ ID NO:30

TABLE 4

Reaction	Primer 1	Primer 2	Approx. Product Size/SEQ ID NO
2.1.1	SEQ ID NO:20	SEQ ID NO:22	none
2.1.2	SEQ ID NO:21	SEQ ID NO:23	899 bp/SEQ ID NO:24
2.2	SEQ ID NO:25	SEQ ID NO:26	846 bp/SEQ ID NO:27
2.3	SEQ ID NO:31	SEQ ID NO:32	424 bp/SEQ ID NO:33
2.4	SEQ ID NO:39	SEQ ID NO:40	460 bp/SEQ ID NO:41
2.5	SEQ ID NO:42	SEQ ID NO:43	235 bp/SEQ ID NO:44

TABLE 5

Reaction	Primer Set PCR 1	Primer Set PCR 2	Approx. Product Size/SEQ ID NO:
3.1	SEQ ID NO:34/SEQ ID NO:35	SEQ ID NO:36/SEQ ID NO:37	1186 bp/SEQ ID NO:38
3.2	SEQ ID NO:45/SEQ ID NO:46	SEQ ID NO:47/SEQ ID NO:48	545 bp/SEQ ID NO:49
3.3	SEQ ID NO:50/SEQ ID NO:51		344 bp/SEQ ID NO:52
3.4	SEQ ID NO:53/SEQ ID NO:54		194 bp/SEQ ID NO:55

TABLE 6

Reaction	Primer Set PCR 1	Primer Set PCR 2	Approx. Product Size/SEQ ID NO:
4.1	SEQ ID NO:56/SEQ ID NO:57	SEQ ID NO:58/SEQ ID NO:59	464 bp/SEQ ID NO:60
4.2	SEQ ID NO:61/SEQ ID NO:62	SEQ ID NO:63/SEQ ID NO:62	433 bp/SEQ ID NO:64
4.3	SEQ ID NO:65/SEQ ID NO:66	SEQ ID NO:65/SEQ ID NO:67	382 bp/SEQ ID NO:68
4.4	SEQ ID NO:69/SEQ ID NO:70	SEQ ID NO:71/SEQ ID NO:72	451 bp/SEQ ID NO:73

TABLE 7

Reaction	Primer Set PCR 1	Primer Set PCR 2	Approx. Product Size/SEQ ID NO:
5.1	SEQ ID NO:74/SEQ ID NO:75	SEQ ID NO:76/SEQ ID NO:77	334 bp/SEQ ID NO:78
5.2	SEQ ID NO:79/SEQ ID NO:80	SEQ ID NO:81/SEQ ID NO:82	413 bp/SEQ ID NO:83

To obtain the sequence at the 3' end of the genome, amplification utilized the 3' RACE System of GIBCO BRL in accordance with the manufacturer's instructions. It was assumed that, as an HEV strain, the 3' end of the HEV-US-1 genome would contain a poly-adenosine tail similar to the Mexican, Burmese, and Pakistani strains. RNA extracted as described above from the equivalent of 50 μ L of serum was reverse transcribed utilizing the oligo dT adapter primer 5'-GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTT-3' of (SEQ ID NO:84) supplied by the manufacturer. First round PCR utilized the AUAP primer supplied 5'-GGCCACGCGTCGACTAGTAC-3' (SEQ ID NO:85) and a HEV US- specific primer (Table 8) at 0.2 mM final concentration with PCR Buffer, $MgCl_2$, and cDNA concentrations as recommended. Amplification involved 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1 minute. Amplification was preceded by a 1 minute incubation at 94°C and followed by a 72°C, 10 minute extension. A second round of amplification used 1 μ L of first round in a 50 μ L reaction. PCR buffer was 1X final concentration with 2 mM $MgCl_2$, and 0.5 mM of each of the primers. Primers were hemi-nested with the AUAP primer and a HEV-US-1 specific primer (Table 8). Amplification conditions were the same as first round. The products were analyzed by agarose gel electrophoresis, cloned, and sequenced as above.

TABLE 8

Reaction	Primer Set PCR 1	Primer Set PCR 2	Approx. Product Size/SEQ ID NO:
8.1	SEQ ID NO:86/SEQ ID NO:85	SEQ ID NO:87/SEQ ID NO:85	960 bp/SEQ ID NO:88

The sequences obtained from the products described in Tables 3, 4, 5, 6, 7, and 8 hereinabove, and the initial PCR product near the 5' end of the genome, SEQ ID NO:15, were assembled into contigs using the programs of the GCG package (Genetics Computer Group, Madison, WI, version 9) and a consensus sequence determined. A schematic of the assembled contig is presented in Figure 3. The HEV US-1 genome is 7202 bp in length, all of which has been sequenced (SEQ ID NO:89). This sequence was translated into three open reading frames, two of which are shown in SEQ ID NO:90 (the third ORF is positioned at nucleotide positions 5094-5462 but cannot be shown in SEQ ID NO:90 due to overlap with the other two ORFs).

The resulting translations (ORF 1, ORF 2, and ORF 3) are set forth in SEQ ID NO:91, SEQ ID NO:92, and SEQ ID NO:93, respectively.

Example 4 - Identification of unique isolate of HEV US-2

A patient from the US suffering from acute hepatitis, who tested for IgG class antibodies in the HEV EIA test, also tested positive by means of a US-1 strain-specific ELISA. This patient (USP-2) diagnosed with acute hepatitis, was a 62 year old male who was admitted to the hospital with jaundice and fatigue. Initial laboratory studies indicated an ALT of 1270 U/L (normal 0-40 U/L). Since there was a recent outbreak of hepatitis A virus (HAV) in the area, it was suspected that this individual was infected with HAV. However, the anti-HAV IgM test, HAVAB-M EIA (Abbott Laboratories) was negative as were tests for serologic markers for hepatitis B virus and hepatitis C virus. This patient's history included a visit to Cancun, Mexico, several weeks prior to the onset of his illness.

The sample from the patient then was analyzed for the presence of HEV specific sequences via PCR amplification using HEV US-1 specific PCR primers. RNA was extracted using Ultraspec as described in Example 2. Random primed cDNA synthesis was performed as described in Example 3 and PCR was performed using standard conditions as described in Example 2 with HEV US-1 specific primers SEQ ID NO:94 and SEQ ID NO:96. Nested PCR was performed with primers SEQ ID NO:95 and SEQ ID NO:97. Sequencing of the PCR product was performed as described in Example 3. The sequence of the resulting PCR product is set forth in SEQ ID NO:98. GAP analysis as described in Example 2 showed that the nucleotide sequence, SEQ ID NO:98 was 95% identical to the corresponding or homologous homologous region from HEV US-1.

Example 5 - Genome Extension and Sequencing of HEV US-2

The clone obtained and sequenced in Example 4 (SEQ ID NO:98) was derived from a HEV isolate most closely related to HEV US-1. To obtain additional regions of the HEV US-2 genome, several RT-PCR walking experiments were performed as described in Example 3.

RNA was extracted using the Total Nucleic Acid Extraction procedure (United States Biochemical). Reverse transcription was random primed using the GeneAmp RNA PCR kit (Perkin-Elmer). Standard PCR was performed in the presence of 2 mM MgCl₂ and 0.5 to 1.0 μM of each primer. Modified reactions contained 1x PCR Buffer and 20% Q Solution (Qiagen) for the isolation of SEQ ID NOS:129, 141 and 146. Reactions used two HEV US-1 specific primers (Table 9), one HEV US-1 specific primer and one HEV consensus primer (Table 10), one HEV US-2 specific primer and one HEV consensus primer (Table 11), two HEV US-2 specific primers (Table 12), or two Burmese, Mexican, and US derived Consensus primers (described hereinbelow, Table 13).

The products shown in SEQ ID NOS:101, 102, 105, 108, 110, 113, 117, 120, 124, 149 and 151 were obtained by touchdown PCR. Amplification involved 43 cycles of 94°C for 30 seconds, 55°C for 30 seconds (-0.3°C/cycle), and 72°C for 1 minute. This was followed by 10 cycles of 94°C for 30 seconds, 40°C for 30 seconds, and 72°C for 1 minute. Cycling involving 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1 minute was used to amplify SEQ ID NOS:129, 132, 136, 141 and 146. All amplifications were preceded by 1-2 minutes at 94°C and followed by 72°C for 5-10 minutes. The reactions were held at 4°C prior to agarose gel analysis. Isolation of many products required a second round of nested or hemi-nested PCR as shown in Tables 9-13. In these reactions 1 μL of the PCR1 product was added to 25-50 μL of the PCR2 reaction mixture and the resulting mixture cycled as in PCR1.

Reactions were analyzed and products cloned and sequenced as described in Example 3 above. The results of these experiments are presented below in Tables 9-13.

TABLE 9

Reaction	Primer set PCR1	Primer set PCR2	Approx. Product Size/SEQ ID NO:
7.1	SEQ ID NO:99/SEQ ID NO:100		331 bp/SEQ ID NO:101
7.2	SEQ ID NO:34/SEQ ID NO:35	SEQ ID NO:36/SEQ ID NO:37	1186 bp/SEQ ID NO:102
7.3	SEQ ID NO:103/SEQ ID NO:104		130bp/SEQ ID NO:105
7.4	SEQ ID NO:106/SEQ ID NO:107	SEQ ID NO:39/SEQ ID NO:107	564 bp/SEQ ID NO:108
7.5	SEQ ID NO: 86/SEQ ID NO:109	SEQ ID NO:87/SEQ ID NO:109	678 bp/SEQ ID NO:110

TABLE 10

Reaction	Primer set PCR1	Primer set PCR2	Approx. Product Size/SEQ ID NO:
8.1	SEQ ID NO:111/SEQ ID NO:112		580 bp/SEQ ID NO:113
8.2	SEQ ID NO:114/SEQ ID NO:116	SEQ ID NO:116/SEQ ID NO:115	734 bp/SEQ ID NO:117

TABLE 11

Reaction	Primer set PCR1	Primer set PCR2	Approx. Product Size/SEQ ID NO:
9.1	SEQ ID NO:118/SEQ ID NO:119		483 bp/SEQ ID NO:120
9.2	SEQ ID NO:121/SEQ ID NO:122	SEQ ID NO:121/SEQ ID NO:123	431 bp/SEQ ID NO:124
9.3	SEQ ID NO:125/SEQ ID NO:126	SEQ ID NO:127/SEQ ID NO:128	1020 bp/SEQ ID NO:129

TABLE 12

Reaction	Primer set PCR1	Primer set PCR2	Approx. Product Size/SEQ ID NO.:
10.1	SEQ ID NO:130/SEQ ID NO:131		407 bp/SEQ ID NO:132
10.2	SEQ ID NO:133/SEQ ID NO:134	SEQ ID NO:135/SEQ ID NO:134	547 bp/SEQ ID NO:136
10.3	SEQ ID NO:137/SEQ ID NO:138	SEQ ID NO:139/SEQ ID NO:140	903 bp/SEQ ID NO:141
10.4	SEQ ID NO:142/SEQ ID NO:143	SEQ ID NO:144/SEQ ID NO:145	503 bp/SEQ ID NO:146

TABLE 13

Reaction	Primer set	Approx. Product Size/SEQ ID NO.:
11.1	SEQ ID NO:147/SEQ ID NO:148	418 bp/SEQ ID NO:149
11.2	SEQ ID NO:150/SEQ ID NO:126	197 bp/SEQ ID NO:151

To obtain the sequence at the 3' end of the genome, amplification utilized the 3' RACE System of GIBCO BRL in accordance with the manufacturer's instructions as described Example 3. cDNA was generated using SEQ ID NO:84. PCR1 utilized primers SEQ ID NO:150 and SEQ ID NO:85. PCR2 primers were SEQ ID NO:152 and SEQ ID NO:85 (reaction 12.1). The resulting product was 901 bp (SEQ ID NO:153).

The isolation of new sequences located at the 5'-terminus of the HEV US-2 viral genome was achieved by inverse PCR (M. Zeiner and U. Gehring, *Biotechniques* 17: 1051-1053, 1994). Due to limited availability of sera from USP-1 and USP-2, fecal material from a HEV US-2 infected macaque (described in Example 9 below) was chosen as the source material. A product of 462 nucleotides was amplified from macaque fecal material from within the hypervariable/ proline rich hinge region using RNA extracted, reverse transcribed, and PCR amplified as described in Example 3 using primers SEQ ID NOS:154, 155, 156 and 157. This product (SEQ ID NO:158) was 100% identical to HEV US-2 sequences. Therefore, it is contemplated that, any sequences identified at the 5' end of the HEV genome from macaque feces should accurately represent the 5' end of the HEV US-2 genome. Total nucleic acids were

extracted from 200 μ L of a 10% fecal suspension as described above. Reverse transcription reactions, which utilized HEV US specific primers (SEQ ID NO:159), were performed using a kit obtained from BMB (as described in M. Zeiner and U. Gehring, *Biotechniques, supra*), except that nucleic acids were denatured at 70°C for 5 min and then placed on ice prior to
5 initiation of the RT reaction. Generation of double-stranded, circular cDNAs was performed as described in M. Zeiner and U. Gehring, *Biotechniques, supra*. The resulting circular cDNA molecules served as template for subsequent PCR reactions. The primers used in the first PCR reaction (PCR1) are shown in SEQ ID NOS:160 and 161. The nested primers used in the second PCR reaction (PCR 2) were as shown in SEQ ID NOS:162 and 163.

10 Products from PCR2 (reaction 13.1) were cloned into pGEM-EasyT Vector (Promega) and sequenced using an Applied Biosystems 373 Automated sequencer. One product of 221 nucleotides was identified as having the appropriate primers and HEV US-2 sequences, identifying 63 nucleotides upstream of known HEV US-2 sequences. Additional clones were identified with the appropriate primers and portions of this new sequence. Primer extension
15 experiments performed on RNA from 100 μ L of USP-2 serum or 100 μ L of a 10% fecal suspension using the sequences shown in SEQ ID NOS:163 and 161 as primers were unsuccessful in confirming the length of this sequence. Pair-wise comparisons of the 63 nucleotides to 5' NTR sequences of Burmese-like isolates revealed identities greater than 94% suggesting that this is the true sequence of HEV US-2.

20 The sequences obtained from the products described in this Example and those described in Example 4 were assembled into contigs using programs in the GCG package (Genetics Computer Group, Madison, WI, version 9) and a consensus sequence determined. A schematic of the assembled contigs is presented in Figure 4. The genome of the HEV US-2 strain is 7277 bp in length, all of which has been sequenced and is set forth in SEQ ID NO:164.
25 This sequence was translated into three open reading frames as indicated in SEQ ID NO:165, with the translation products of the ORF 1 and ORF 2 sequences only being shown (the third ORF is positioned at nucleotide positions 5159-5527 but cannot be shown within SEQ ID NO:165 due to overlap with the other two ORFs). The resulting translations of the ORF 1, ORF 2, and ORF 3 sequences are shown in SEQ ID NOS:166, 167 and 168, respectively.

Example 6 - Sequence Comparisons

Information about the degree of relatedness of viruses typically can be obtained by performing comparisons such as alignments of nucleotide and deduced amino acid sequences. Alignments of the sequences of the US isolates of HEV (*e.g.*, HEV US-1 and HEV US-2) with corresponding sequences of other isolates of HEV provide a quantitative assessment of the degree of similarity and identity between the sequences. In general, the calculation of the similarity between two amino acid sequences is based upon the degree of likeness exhibited between the side chains of an amino acid pair in an alignment. The degree of likeness is based upon the physical-chemical characteristics of the amino acid side chains, *i.e.* size, shape, charge, hydrogen-bonding capacity, and chemical reactivity. Thus similar amino acids possess side chains that have similar physical-chemical characteristics. The calculation of identity between two aligned amino acid or nucleotide sequences is, in general, an arithmetic calculation that counts the number of identical pairs of amino acids or nucleotides in an alignment and divides this number by the length of the sequence(s) in the alignment. The calculation of similarity between two aligned nucleotide sequences sometimes uses different values for transitions and transversions between paired (*i.e.* matched) nucleotides at various positions in the alignment. However, the magnitude of the similarity and identity scores between pairs of nucleotide sequences, are usually very close, *i.e.* within one to two percent.

The degree of similarity and identity was determined using the program GAP of the Wisconsin Sequence Analysis Package (Version 9). The gap creation and gap extension penalties were 50 and 3.0, respectively, for nucleic acid sequence alignments, and 12 and 4, respectively, for amino acid sequence comparisons.

As indicated previously, a partial identity exists between the initial 5'-end ORF 1 clone and other isolates of HEV, which supports the proposition that the HEV infection associated with patient USP-1 is due to a unique isolate of HEV. In order to more extensively determine the degree of relatedness between this isolate and other known isolates of HEV, alignments of the extended nucleotide and deduced amino acid sequences were performed.

Pair-wise nucleotide and amino acid comparisons of HEV US-1, HEV US-2, and 10 other full length HEV genomes (obtained from a publicly-available database, see Table 14) were performed, as described above, to determine the relationship of the US isolates to each other and to the known variants of HEV.

TABLE 14

Isolate	Genbank Accession Number
Mexican (M1)	M74560
Burmese (B1)	M73218
Burmese (B2)	D10330
Pakistan (P1)	M80581
Chinese (C1)	D11092
Chinese (C2)	L25547
Chinese (C3)	M94177
Chinese (C4)	D11093
Indian (I1)	X98292
Indian (I2)	X99441

5 Nucleotide identity across the entire genomes of US-1, US-2, B1, B2, I2, C1, C2, C3, P1, C4 and I1 strains is presented in Table 15. The nucleotide identities of ORF 1, ORF 2, and ORF 3 are shown in Tables 16, 17 and 18, respectively. Tables 17 and 18 also contain comparisons against a recently isolated swine (S1) sequence, available under GenBank accession number AF011921.

TABLE 15 - Nucleotide Identity Across Genome

	US-1	US-2	B1	B2	I2	C1	C2	C3	P1	C4	I1
US-2	92.0										
B1	73.9	74.0									
B2	73.8	74.0	98.5								
I2	73.5	73.8	96.1	95.4							
C1	74.2	74.3	93.9	93.4	92.3						
C2	74.2	74.3	93.5	93.0	92.0	98.7					
C3	74.1	74.3	93.7	93.0	92.0	98.2	98.7				
P1	74.1	74.1	93.6	92.8	92.0	98.2	98.8	98.3			
C4	73.7	73.9	94.5	94.1	92.7	97.1	97.2	96.8	96.7		
I1	74.4	74.4	93.5	93.0	92.2	93.8	94.0	93.8	93.9	93.5	
M1	73.7	74.5	75.9	75.7	75.0	75.9	75.9	75.9	76.1	75.7	75.7

TABLE 16 - Nucleotide Identity Across ORF 1

	US-1	US-2	B1	B2	I2	C1	C2	C3	P1	C4	I1
US-1											
US-2	92.0										
B1	71.7	71.6									
B2	71.7	71.8	98.6								
I2	71.2	71.5	95.7	95.1							
C1	72.1	72.1	93.5	93.1	91.8						
C2	72.2	72.3	93.1	92.7	91.5	98.6					
C3	71.9	72.2	93.3	92.8	91.4	98.1	98.7				
P1	72.2	72.1	93.1	92.6	91.4	98.2	99.0	98.4			
C4	71.5	71.7	94.6	94.4	92.3	96.7	98.8	96.3	96.4		
I1	72.3	72.3	93.2	92.8	91.5	93.6	94.0	93.7	93.9	93.3	
M1	72.0	72.6	73.6	73.5	72.5	73.7	73.8	73.8	73.9	73.4	73.5

TABLE 17 - Nucleotide Identity Across ORF 2

	US-1	US-2	B1	B2	I2	C1	C2	C3	P1	C4	I1	M1
US-1												
US-2	92.2											
B1	79.2	79.6										
B2	86.4	79.4	98.5									
I2	79.0	79.5	99.2	98.4								
C1	79.3	79.5	94.4	98.4	98.4							
C2	79.2	79.4	94.3	97.8	97.8	98.9						
C3	79.3	79.4	94.4	97.8	97.8	98.9	98.4					
P1	79.0	79.3	93.8	98.1	98.7	99.7	99.2	99.2				
C4	78.8	79.3	94.0	97.8	97.8	98.9	98.4	98.4	97.4			
I1	79.4	79.7	94.1	97.6	97.3	97.9	97.0	94.0	93.7	93.9		
M1	78.0	79.3	81.1	90.1	98.5	90.6	90.1	81.0	81.4	90.3	90.3	
S1	92.0	98.9	79.8	84.6	85.4	85.4	85.1	80.2	80.1	84.8	85.1	84.6

TABLE 18 - Nucleotide Identity Across ORF 3

	US-1	US-2	B1	B2	I2	C1	C2	C3	P1	C4	I1	M1
US-1												
US-2	96.2											
B1	87.0	86.6										
B2	86.4	86.3	99.2									
I2	86.4	86.9	97.8	99.2								
C1	87.3	86.3	99.2	98.4	98.4							
C2	86.4	86.1	98.1	97.3	97.8	98.9						
C3	86.7	85.6	98.1	97.3	97.8	98.9	98.4					
P1	87.0	86.6	98.9	98.1	98.7	99.7	99.2	99.2				
C4	86.2	85.8	98.1	97.6	97.8	98.9	98.4	98.4	99.2			
I1	86.4	86.6	97.8	97.6	97.6	97.9	97.0	97.0	97.8	97.8		
M1	84.6	85.2	87.8	90.1	89.5	90.6	90.1	90.1	90.9	90.3	90.3	
S1	94.9	96.7	85.1	84.6	85.4	85.4	85.1	84.8	85.6	84.8	85.1	84.6

In addition, the ORF 1 nucleotide sequences encoding the methyltransferase proteins were compared between each of the US-1, US-2, M1 and P1 isolates. The methyltransferase encoding region of the HEV US-1 genome is represented by residues 1-693 of SEQ ID NO:89, whereas the methyltransferase encoding region of the HEV US-2 genome is represented by residues 36-755 of SEQ ID NO:164. The comparison results are set forth in Table 19.

TABLE 19 - Methyltransferase Region

	% IDENTITY			
	US-1	US-2	M1	P1
US-1	-	93.4	77.0	75.2
US-2	-	-	78.5	76.0
M1	-	-	-	78.8

The ORF 1 nucleotide sequences encoding the Y domain proteins were compared between each of the US-1, US-2, M1 and P1 isolates. The Y domain protein encoding region of the HEV US-1 genome is represented by residues 619-1272 of SEQ ID NO:89, whereas the Y domain protein encoding region of the HEV US-2 genome is represented by residues 680-1334 of SEQ ID NO:164. The comparison results are set forth in Table 20.

TABLE 20 - Y Domain

	% IDENTITY			
	US-1	US-2	M1	P1
US-1	-	94.0	79.0	77.2
US-2	-	-	79.7	76.8
M1	-	-	-	78.3

The ORF 1 nucleotide sequences encoding the protease proteins were compared between each of the US-1, US-2, M1 and P1 isolates. The protease protein encoding region of the HEV US-1 genome is represented by residues 1270-2091 of SEQ ID NO:89, whereas the protease protein encoding region of the HEV US-2 genome is represented by residues 1332-2153 of SEQ ID NO:164. The comparison results are set forth in Table 21.

TABLE 21 - Protease Region

	% IDENTITY			
	US-1	US-2	M1	P1
US-1	-	91.8	65.1	64.0
US-2	-	-	65.1	63.1
M1	-	-	-	68.1

The ORF 1 nucleotide sequences encoding the hypervariable region were compared between each of the US-1, US-2, M1 and P1 isolates. The hypervariable region encoding region of the HEV US-1 genome is represented by residues 2092-2364 of SEQ IS NO:89, whereas the hypervariable region encoding region of the HEV US-2 genome is represented by residues 2194-2429 of SEQ ID NO:164. The comparison results are set forth in Table 22.

TABLE 22 - Hypervariable Region

	% IDENTITY			
	US-1	US-2	M1	P1
US-1	-	83.9	40.3	50.2
US-2	-	-	45.8	49.8
M1	-	-	-	40.4

The ORF 1 nucleotide sequences encoding the X domain proteins were compared between each of the US-1, US-2, M1 and P1 isolates. The X domain protein encoding region of the HEV US-1 genomes represented by residues 2365-2841 of SEQ ID NO:89, whereas the X domain probe encoding region of the HEV US-2 genome is represented by residues 2430-2906 of SEQ ID NO:164. The comparison results are set forth in Table 23.

TABLE 23 - X Domain

	% IDENTITY			
	US-1	US-2	M1	P1
US-1	-	91.6	72.5	71.3
US-2	-	-	72.7	70.9
M1	-	-	-	72.9

The ORF 1 nucleotide sequences encoding the helicase proteins were compared between each of the US-1, US-2, M1 and P1 isolates. The helicase encoding region of the HEV US-1 genomes represented by residues 2893-3591 of SEQ ID NO:89, whereas the helicase encoding region of the HEV US-2 genome is represented by residues 2958-3656 of SEQ ID NO:164. The comparison results are set forth in Table 24.

TABLE 24 - Helicase Region

	% IDENTITY			
	US-1	US-2	M1	P1
US-1	-	92.8	76.5	75.2
US-2	-	-	75.4	74.1
M1	-	-	-	76.2

The ORF 1 nucleotide sequences encoding the RNA-dependent RNA polymerase proteins were compared between each of the US-1, US-2, M1 and P1 isolates. The polymerase encoding region of the HEV US-1 genome is represented by residues 3634-5094 of SEQ ID NO:89, whereas the polymerase encoding region of the HEV US-2 genome is represented by residues 3699-5159 of SEQ ID NO:164. The comparison results are set forth in Table 25.

TABLE 25 - RNA-dependent RNA Polymerase Region

	% IDENTITY			
	US-1	US-2	M1	P1
US-1	-	93.1	72.9	75.3
US-2	-	-	73.6	75.8
M1	-	-	-	77.1

In addition, the amino acid identities/similarities of the proteins encoded by the ORF 1, ORF 2, and ORF 3 sequences of US-1, US-2, B1, B2, I2, C1, C2, C3, P1, C4 and I1 strains are shown in Tables 26, 27 and 28 respectively. In addition, Tables 27 and 28 also contain comparisons against the swine sequence (S1). In Tables 26, 27 and 28, the similarities are presented in the upper right hand halves of the tables and the identities are presented in the lower left hand halves of the tables.

TABLE 26 - Amino Acid Similarity/Identity Across ORF 1

% I D E N T I T Y	% SIMILARITY												
		US-1	US-2	B1	B2	I2	C1	C2	C3	P1	C4	I1	M1
	US-1		97.8	86.0	85.7	84.4	85.9	86.2	84.9	86.4	85.7	86.3	85.4
	US-2	97.5		86.2	85.8	84.5	85.8	86.0	85.0	86.3	85.7	86.3	85.5
	B1	82.4	82.6		98.7	96.8	98.4	98.5	97.1	98.5	98.1	98.2	87.0
	B2	82.3	82.3	98.6		96.2	97.8	97.9	96.3	97.8	97.6	97.6	86.6
	I2	80.7	80.7	96.3	95.7		96.3	96.4	95.0	96.3	95.9	95.9	85.2
	C1	82.5	82.3	98.2	97.5	95.7		99.5	97.9	99.4	99.0	98.2	86.9
	C2	82.8	82.6	98.4	97.8	95.9	99.4		98.2	99.6	99.2	98.4	87.0
	C3	81.6	81.6	96.9	96.1	94.4	97.7	98.1		98.1	97.6	97.0	85.9
	P1	83.0	82.9	98.4	97.7	95.9	99.2	99.6	98.0		99.0	98.4	87.1
	C4	82.5	82.3	98.0	97.6	95.4	98.8	99.1	97.4	98.9		97.8	86.5
	I1	82.9	82.9	98.1	97.5	95.5	98.1	98.4	96.9	98.4	97.8		87.3
	M1	82.0	82.0	83.8	83.4	81.8	83.7	83.9	82.8	84.0	83.4	84.2	

TABLE 27 - Amino Acid Similarity/Identity Across ORF 2

% I D E N T I T Y	% SIMILARITY													
		US-1	US-2	B1	B2	I2	C1	C2	C3	P1	C4	I1	M1	S1
	US-1		98.3	93.3	93.0	93.0	93.5	93.2	92.9	93.2	92.4	92.6	91.5	97.1
	US-2	98.0		93.3	93.0	93.3	93.3	93.3	93.0	93.3	92.6	92.7	91.7	99.1
	B1	91.8	91.8		98.9	99.1	99.8	99.2	99.2	99.5	98.8	98.9	94.8	93.0
	B2	91.5	91.5	98.9		98.3	99.1	98.5	98.5	98.8	98.2	98.2	94.1	92.7
	I2	91.5	91.8	99.1	98.3		99.2	98.9	98.6	99.2	98.5	98.6	94.5	91.5
	C1	92.0	92.0	99.7	98.9	99.1		99.4	99.1	99.7	98.9	99.1	95.0	93.2
	C2	91.7	92.0	99.1	98.3	98.8	99.4		98.8	99.4	98.6	98.8	94.7	93.0
	C3	91.4	91.7	99.1	98.3	98.5	99.1	98.8		99.1	98.3	98.5	94.4	92.7
	P1	91.7	92.0	99.4	98.6	99.1	99.7	99.4	99.1		98.9	99.1	95.0	93.0
	C4	90.9	91.2	98.6	98.0	98.4	98.9	98.6	98.3	98.9		98.3	94.2	92.3
	I1	91.1	91.4	98.5	97.7	98.2	98.8	98.5	98.2	98.8	98.0		94.7	92.4
	M1	90.1	90.6	93.2	92.4	92.9	93.3	93.0	92.9	93.3	92.6	93.0		91.2
	S1	97.7	98.9	91.7	91.4	91.9	91.8	91.7	91.4	91.7	90.9	91.1	90.2	

TABLE 28 - Amino Acid Similarity/Identity Across ORF 3

		% SIMILARITY												
%		US-1	US-2	B1	B2	I2	C1	C2	C3	P1	C4	I1	M1	S1
	US-1		96.7	85.2	84.4	85.2	85.2	83.6	85.2	85.2	83.6	85.2	79.5	93.5
	US-2	96.7		85.2	84.4	85.2	85.2	83.6	83.6	85.2	83.6	85.2	81.1	96.7
	I	B1	84.4	84.4		98.4	100.0	100.0	98.4	98.4	100.0	98.4	87.0	83.7
	D	B2	83.6	83.6	98.4		98.4	98.4	96.7	96.7	98.4	96.7	87.0	82.9
	E	I2	84.4	84.4	100.0	98.4		100.0	98.4	98.4	100.0	98.4	87.0	83.7
	N	C1	84.4	84.4	100.0	98.4	100.0		98.4	98.4	100.0	98.4	87.0	83.7
	T	C2	82.8	82.8	98.4	96.7	98.4	98.4		96.7	98.4	97.6	96.7	85.4
	I	C3	84.4	82.8	98.4	96.7	98.4	98.4	96.7		98.4	96.7	96.7	85.4
	T	P1	84.4	84.4	100.0	98.4	100.0	100.0	98.4	98.4		98.4	98.4	87.0
	Y	C4	82.8	82.8	98.4	96.7	98.4	98.4	97.6	96.7	98.4		96.7	85.4
		I1	84.4	84.4	98.4	96.7	98.4	98.4	96.7	96.7	98.4	96.7		88.6
		M1	78.7	80.3	87.0	87.0	87.0	87.0	85.4	85.4	87.0	85.4	88.6	
		S1	93.5	96.7	82.9	82.1	82.9	82.9	81.3	81.3	82.9	81.3	82.9	78.9

In addition, the ORF 1 amino acid sequences defining the methyltransferase proteins were compared between each of the US-1, US-2, M1 and P1 isolates. The methyltransferase protein encoded by the HEV US-1 genome is represented by residues 1-231 of SEQ ID NO:91, whereas the methyltransferase protein encoded by the HEV US-2 genome is represented by residues 1-240 of SEQ ID NO:166. The comparison results are set forth in Table 29.

TABLE 29 - Methyltransferase Region

		% IDENTITY			
%		US-1	US-2	M1	P1
	US-1	-	98.7	91.3	88.7
	US-2	98.7	-	91.7	89.1
	M1	91.8	92.0	-	92.9
	P1	90.0	90.4	91.2	-

The ORF 1 amino acid sequences defining the protease proteins were compared between each of the US-1, US-2, M1 and P1 isolates. The protease protein encoded by the

HEV US-1 genome is represented by residues 424-697 of SEQ ID NO:91, whereas the protease protein encoded by the HEV US-2 genome is represented by residues 433-706 of SEQ ID NO:166. The comparison results are set forth in Table 30.

TABLE 30 - Protease Region

% S I M I L A R I T Y	% IDENTITY				
		US-1	US-2	M1	P1
	US-1	-	98.5	67.5	69.3
	US-2	97.8	-	67.1	68.6
	M1	73.3	73.3	-	76.6
	P1	74.4	74.0	72.2	-

The ORF 1 amino acid sequences defining Y domain proteins were compared between each of the US-1, US-2, M1 and P1 isolates. The Y domain protein encoded by the HEV US-1 genome is represented by residues 207-424 of SEQ ID NO:91, whereas the Y domain protein encoded by the HEV US-2 genome is represented by residues 216-433 of SEQ ID NO:166. The comparison results are set forth in Table 31.

TABLE 31 - Y Domain

% S I M I L A R I T Y	% IDENTITY				
		US-1	US-2	M1	P1
	US-1	-	98.2	92.7	93.6
	US-2	98.2	-	92.7	93.6
	M1	94.0	94.0	-	93.1
	P1	94.5	94.5	91.7	-

The ORF 1 amino acid sequences defining the X domain proteins were compared between each of the US-1, US-2, M1 and P1 isolates. The X domain encoded by the HEV US-1 genome is represented by residues 789-947 of SEQ ID NO:91, whereas the X domain protein encoded by the HEV US-2 genome is represented by residues 799-957 of SEQ ID NO:166.

5 The comparison results are set forth in Table 32.

TABLE 32 - X Domain

% S I M I L A R I T Y	% IDENTITY				
		US-1	US-2	M1	P1
	US-1	-	97.5	82.4	80.5
	US-2	97.5	-	81.8	79.9
	M1	88.0	87.4	-	86.1
	P1	84.3	83.6	83.0	-

The ORF 1 amino acid sequences defining helicase proteins were compared between each of the US-1, US-2, M1 and P1 isolates. The helicase encoded by the HEV US-1, US-2, M1 and P1 isolates. The helicase encoded by the HEV US-1 genome is represented by residues 965-1197 of SEQ ID NO:91, whereas the helicase encoded by the HEV US-2 genome is

represented by residues 975-1207 of SEQ ID NO:166. The comparison results are set forth in Table 33.

TABLE 33 - Helicase Region

% S I M I L A R I T Y	% IDENTITY				
		US-1	US-2	M1	P1
	US-1	-	99.1	89.7	91.0
	US-2	99.1	-	90.6	91.8
	M1	93.1	94.0	-	95.2
	P1	94.0	94.8	91.0	-

The ORF 1 amino acid sequence defining the hypervariable regions were compared between each end of the US-1, US-2, M1 and P1 isolates. The hypervariable region encoded by the HEV US-1 genome is represented by residues 698-788 of SEQ ID NO:91, whereas the hypervariable region encoded by the HEV US-2 genome is represented by residues 707-798 of SEQ ID NO:166. The comparison results are set forth in Table 34.

TABLE 34 - Hypervariable Region

% S I M I L A R I T Y	% IDENTITY				
		US-1	US-2	M1	P1
	US-1	-	82.4	25.0	27.7
	US-2	79.1	-	25.0	21.0
	M1	25.0	25.0	-	20.8
	P1	31.9	21.0	18.0	-

The ORF 1 amino acid sequence defining the RNA-dependent RNA polymerase proteins were compared between each of the US-1, US-2, M1 and P1 isolates. The polymerase encoded by the HEV US-1 genome is represented by residues 1212-1698 of SEQ ID NO:91, whereas the polymerase encoded by the HEV US-2 genome is represented by residues 1222-1708 of SEQ ID NO:166. The comparison results are set forth in Table 35.

TABLE 35 - RNA-dependent RNA Polymerase Domain

% S I M I L A R I T Y	% IDENTITY				
		US-1	US-2	M1	P1
	US-1	-	99.0	86.0	87.8
	US-2	99.0	-	86.2	87.7
	M1	89.7	89.9	-	92.6
	P1	91.6	91.6	89.5	-

In addition to the foregoing, several additional HEV isolates belonging to the HEV US-type family were identified during the course of this work (see, Example 13 below). The additional isolates are denoted as It1 (Italian strain), G1 (first Greek strain) and G2 (second Greek strain). Additional sequence comparisons were performed and include the It1, G1 and G2 sequences, the results of which are presented below in Tables 36 and 37. Table 36 shows the nucleotide and deduced amino acid identities between isolates of HEV over a 371 base (123 amino acids) ORF 1 fragment. The ORF 1 fragment corresponds to residues 26-396 of SEQ ID NO:89. Table 37 shows the nucleotide and deduced amino acid identities between isolates of HEV over a 148 base (49 amino acid) ORF 2 fragment. The ORF 2 fragment corresponds to residues 6307-6454 of SEQ ID NO:89. In both Tables 36 and 37, the isolates represented are Burmese (B1, B2), Chinese (C1, C2, C3, C4), Indian (I1, I2), Pakistan (P1), Mexican (M1), Swine (S1), United States (US-1, US-2), Greek (G1, G2) and Italian (It1).

Pairwise comparisons of the full length nucleotide sequences were preferred using the nucleotide sequences of the respective genomes of HEV US-1 and HEV US-2 together with the other genomes of the other HEV isolates identified in Table 14. The results of the comparison are shown in Table 15. At the nucleotide level, HEV US-1 and HEV US-2 were most closely related to each other, with 92.0% identity across the entire genome. The full length Burmese-like isolates demonstrated similar identities ranging from 92.0 to 98.8%. The US isolates were 73.5 to 74.5% identical to the Burmese-like and Mexican isolates. This is similar to the identity seen between any one Burmese-like isolate and the Mexican isolate, 75.0 to 76.1% nucleotide identity. These data indicate that the US isolates are members of a new strain variant of HEV, distinct from the Burmese and Mexican strains.

Similar degrees of identity are found when smaller portions of each genome are analyzed, such as the individual ORFs. These values are presented in Tables 16, 17 and 18 for ORF 1, ORF 2, and ORF 3, respectively. Across each region, the Burmese and Pakistani isolates demonstrate the highest degree of identity ranging from 93.1 to 98.9% identity. The Mexican isolate is distinct, with identities of 73.6 to 90.1% to the Burmese-like isolates. HEV US-1 nucleotide sequence analysis reveals a significant degree of divergence with ORF 1 sequences being less than 72% identical to the Burmese-like and Mexican isolates. Similarly, ORF 2 and ORF 3 sequences were less than 79.1% and 86.9% identical to the Burmese-like and Mexican isolates, respectively.

The variability seen at the nucleotide level is reflected in the amino acid similarity and identity of the translated open reading frames. ORF 1 is the most divergent product, potentially due to the presence of a hypervariable region. The US isolates possess 97.5% amino acid identity across this region (Table 26). This is similar to the 94.4 to 99.6% identity seen between Burmese-like ORF 1 proteins. The US ORF 1 products are 80.7 to 83.0% identical to Burmese-like and Mexican proteins (Table 26). These values are similar to those observed between any one Burmese-like isolates and the Mexican isolate, ranging from 81.8 to 84.2% identity. Amino acid similarity values are generally up to 3.5% higher than the identity value, reflecting a large number of conservative amino acid substitutions. The ORF 2 product is the most conserved, potentially due to its role as the viral capsid protein. The US ORF 2 products

are 98.0% identical to each other, while being 90.1 to 92% identical to Burmese and Mexican ORF 2 proteins (Table 27). Again, these ranges mirror those observed between Burmese isolates (97.7 to 99.7% identity). Identity between Burmese and Mexican isolates is slightly greater than that between the US variant and other variants, being 92.4 to 93.3%. Amino acid similarity across ORF 2 adds approximately 1.5% to the identity value. The ORF 3 product of HEV US-1 and HEV US-2 shared 96.7% amino acid identity. The Burmese isolates showed 96.7 to 100% amino acid identity. ORF 3 amino acid identities of the US isolates to the Burmese and Mexican isolates were 78.7 to 84.4%, slightly less than that observed between Burmese and Mexican isolates, 85.4 to 88.6% identity (Table 28). Amino acid similarity across ORF 3 was generally the same as the identity values, however, some comparisons demonstrated similarity values less than 1.0% greater than the identity value. These amino acid similarity and identity values indicate that the analysis of short amino acid sequences produce similar results to full length and partial nucleotide analyses, indicating that the US isolates are closely related and genetically distinct from previously characterized isolates of HEV.

Tables 27 and 28 also include pairwise amino acid sequence comparisons with a HEV-like isolate recently identified in swine (Meng *et al.* (1997) Proc. Natl. Acad. Sci. USA 94: 9860-9865. Only 2021 bp across the ORF 2/3 region have been characterized (GenBank Accession Number: AF011921). The US swine sequence is 92% identical to the corresponding region of HEV US-1 at the nucleotide level. It is noted that HEV US-1 is very similar at the amino acid level to the recently identified swine virus. For example, the HEV US-1 and swine strains exhibit 97.1% and 93.5% identity over the respective ORF 2 and ORF 3 sequences (Tables 27 and 28, respectively).

Partial sequences of 210 nucleotides from two HEV isolates from China referred to as G9 and G20 (Genbank Accession numbers X87306 and X87307, respectively) recently have been described in the literature by (Huang *et al.* (1995) J. Med Virology 47: 303-308). These fragments represent nucleotide sequences homologous to residue numbers 4533 to 4742 of SEQ ID NO:89. Their encoded amino acid sequences (69 amino acid residues in length) are homologous to residue numbers 1512-1580 of SEQ ID NO:91. The results from the pairwise comparisons of the nucleotide sequences and the predicted amino acid sequences of these

sequences are shown in Tables 38 and 39. Results indicate that the G9 and G20 isolates are 89% identical to one another at the nucleotide level across this region. The closely related Burmese and Pakistan isolates are 92.9% identical over this range. The US-1 isolate exhibits a 77.1 and 81.0 across this region suggesting that the US-1 isolate also is unique from these isolates. Although the G9 and G20 sequences are most closely related at the nucleotide level, the deduced amino acid translation of G20 is most similar/identical to the US sequence from the US-1 isolate (Table 38). This is most likely due to the short length of amino acids utilized in the analysis.

TABLE 38. Identity across 210 nucleotides of ORF 1

	Pak	Mex	US-1	G20	G9
Bur	92.9	74.8	75.7	78.1	76.7
Pak		75.2	76.7	78.1	76.7
Mex			77.1	75.2	71.9
US-1				81.0	77.1
G20					89.0

TABLE 39. Similarity/identity across 69 amino acids of ORF 1

	Pak	Mex	US-1	G20	G9
Bur	98.6 / 98.6	92.8 / 88.4	92.8 / 85.5	92.8 / 88.4	82.6 / 79.7
Pak		94.2 / 89.9	91.3 / 84.1	91.3 / 87.0	84.1 / 81.2
Mex			89.9 / 87.0	89.9 / 87.0	81.2 / 78.3
US-1				100 / 95.7	88.4 / 88.1
G20					88.4 / 87.0

Example 7 - Phylogenetic Analyses.

Alignments of nucleotide and amino acid sequences were performed in order to determine the phylogenetic relationships between the novel US-type isolates and other isolates of HEV. The alignments were made using the program PILEUP of the Wisconsin Sequence Analysis Package, version 9 (Genetics Computer Group, Madison, WI). Evolutionary distances between sequences were determined using the DNADIST program (Kimura 2-parameter method) with a transition-transversion ratio of 2.0 and PROTDIST (Dayhoff PAM matrix) program of the PHYLIP package, version 3.5c (Felsenstein 1993, Department of Genetics, University of Washington, Seattle). The computed distances were used for the construction of phylogenetic trees using the program FITCH (Fitch-Margoliash method). The robustness of the trees was determined by bootstrap resampling of the multiple-sequence alignments (100 sets or 1,000 sets) with the programs SEQBOOT, DNADIST, the neighbor-joining method of the program NEIGHBOR, and CONSENSE (PHYLIP package). Bootstrap values of less than 70% are regarded as not providing evidence for a phylogenetic grouping (Muerhoff *et al.*, (1997) Journal of Virology, 71: 6501-6508). The final trees were produced using RETREE (PHYLIP) with the midpoint rooting option and the graphical output was created with TREEVIEW (Page, (1996) Computer Applied Biosciences 12: 357-358), the results of which are presented in Figures 5, 6, 10, and 11.

Phylogenetic analysis with complete genomes. To more extensively determine the degree of relatedness between HEV US-1, HEV US-2, and other known isolates of HEV, nucleotide alignments were performed. The full length HEV US-1 and HEV US-2 genomes were aligned with 10 other isolates of HEV from which complete genomes are available (Table 14).

Examination of the phylogenetic distances based upon alignments of the HEV-US isolates and other isolates of HEV demonstrate that there is considerable evolutionary distance between those from the US and those from other geographical areas as determined using the DNADIST program (Kimura 2-parameter method) with a transition-transversion ratio of 2.0 (Table 40). The distances calculated also show the close relationship between the isolates originating from Asia. Within this Burmese-like group the maximum distance calculated from the full length alignment is 0.0850 nucleotide substitutions per base. The minimum distance between a

member of this group and a US isolate is 0.3322 substitutions. The Mexican strain shows similar distances to the Burmese-like group of 0.3055 to 0.3132 substitutions and 0.3322 to 0.3462 substitutions to the US isolate. The genetic distance between HEV US-1 and HEV US-2 of 0.0812 substitutions is similar to that seen between Burmese-like isolates. The relative evolutionary distances between the viral sequences analyzed are readily apparent upon inspection of the unrooted phylogenetic tree presented in Figure 5, where the branch lengths are proportional to the evolutionary distances. In the phylogenetic tree, the Burmese-like isolates, the Mexican isolate and the US isolates each represent a major branch. In addition, the branching of the prototype viruses are supported with bootstrap values of 100%. Analysis of smaller segments of the genome (e.g. ORF 1, ORF 2, or ORF 3) were individually analyzed resulting in trees analogous to those obtained with the full length sequence and shown in Figure 5. These analyses demonstrate that the HEV US isolates represent a distinct strain or variant of HEV and that HEV US-1 and HEV US-2 are as similar to each other as are the most divergent Burmese-like isolates.

TABLE 40 - Phylogenetic distances over the full length sequence

	B1	B2	C1	C2	C3	C4	I1	I2	P1	M1	US-1
B1											
B2	0.0149										
C1	0.0643	0.0697									
C2	0.0680	0.0733	0.0136								
C3	0.0663	0.0734	0.0178	0.0132							
C4	0.0574	0.0611	0.0304	0.0290	0.0329						
I1	0.0677	0.0728	0.0645	0.0625	0.0647	0.0681					
I2	0.0403	0.0477	0.0820	0.0849	0.0846	0.0776	0.0832				
P1	0.0693	0.0751	0.0178	0.0120	0.0172	0.0335	0.0633	0.0850			
M1	0.3096	0.3120	0.3086	0.3089	0.3091	0.3132	0.3120	0.3259	0.3055		
US-1	0.3406	0.3418	0.3360	0.3345	0.3367	0.3445	0.3322	0.3464	0.3363	0.3462	
US-2	0.3413	0.3408	0.3370	0.3361	0.3374	0.3445	0.3333	0.3461	0.3377	0.3367	0.0812

Comparison to ORF 2/ORF 3 from Swine HEV. In order to determine the relationship between a recently described swine-HEV and the human HEV US-1 and HEV US-2 isolates, comparisons of the nucleotide sequences across the complete ORF 2 and ORF 3 were performed using analogous regions from the 10 full length sequences utilized above (Table 14).

- 5 Phylogenetic analysis produces genetic distances of 0.0799 to 0.0810 nucleotide substitutions per position between the US and swine HEV isolates (Table 41). These values are similar to those observed between the most distant Burmese-like isolates. The US and swine isolates group closely on an unrooted phylogenetic tree when the ORF 2/3 nucleotide sequences are analyzed (See, Figure 6). These isolates form a phylogenetic group distinct from the Mexican
- 10 isolate and the Burmese-like isolates. These grouping are supported by bootstrap values of 100%.

TABLE 41 - Phylogenetic distances between USswine and human HEV isolates

	US-2	USswine	Burmese	Mexican
US-1	0.0799	0.0810	0.2441-0.2495	0.2671
US-2		0.0795	0.2409-0.2479	0.2486
USswine			0.2348-0.2485	0.2615
Burmese			0.0119-0.0716	0.2183-0.2248

Example 8 - HEV Serologic Studies

A. Background

Early studies indicate that epitopes useful for diagnosis of HEV infections are located near the carboxyl terminus of ORF 2 and ORF 3 of both the Burmese and Mexican strains of HEV. The two antigens from the Mexican strain, referred to hereinafter as M 3-2 and M 4-2, comprise 42 and 32 amino acids near the carboxyl terminus of ORF 2 and ORF 3, respectively (Yarbough *et al.* (1991) *Journal of Virology*, 65: 5790-5797). The two antigens from the Burmese strain of HEV, referred to hereinafter as B 3-2 and B 4-2 proteins, comprise 42 and 33 amino acids near the carboxyl terminus of ORF 2 and ORF 3, respectively (Yarbough *et al.* (1991) *supra*). Diagnostic tests designed to detect IgG, IgA and IgM class antibodies to HEV have been developed based on these antigenic regions. Additional HEV recombinant proteins have been generated that encompass full-length ORF 3 (Dawson *et al.* (1992) *Journal of Virology Methods*, 38: 175-186) or additional amino acid sequences from the ORF 2 protein (Dawson *et al.* (1993) *supra*), to potentially enhance the detection of antibodies to HEV. Comparative studies indicate that the original recombinant proteins and synthetic peptides (B4-2, B3-2, M3-2, M4-2) were as effective as the larger recombinant proteins in detecting antibodies to HEV in known cases of acute HEV infection. A licensed test to detect antibodies to HEV is manufactured by Abbott Laboratories and consists of the full length Burmese strain ORF 3 protein and the carboxyl 327 amino acids of the Burmese strain ORF 2 protein.

After initial serological studies demonstrating the utility of B 3-2, B 4-2, M 3-2 and M

4-2, it was established that six additional amino acids reside at the carboxyl terminus of ORF 2 of both the Burmese and Mexican strains of HEV which do not form part of the M 3-2 and B 3-2 antigenic peptides. Since the carboxyl ends of ORF 2 and ORF 3 have been shown to be of value for the Burmese and Mexican strains of HEV, synthetic peptides corresponding to the

5 these regions of the genome were generated for the US-1 strain of HEV. The synthetic peptides corresponding to the 48 amino acids at the carboxyl end of the ORF 2 were generated for the Burmese and Mexican strains of HEV (SEQ ID NOS:172 and 170, respectively), and are referred to as B 3-2e and M 3-2e (where "e" designates extended amino acid sequence). In addition, synthetic peptides representing the 33 amino acids at the carboxyl end of the HEV

10 US-1 ORF 3 were generated for the Burmese and Mexican strains of HEV (SEQ ID NOS:171 and 169, respectively), and are referred to as B4-2 and M4-2. The synthetic peptide based on the epitope from within ORF 2 for the HEV US-1 strain (SEQ ID NO:174) is referred to as the US 3-2e. The synthetic peptide based on the epitope at the carboxyl end of the HEV US-1 ORF 3 (SEQ ID NO:173) is referred to as US 4-2. Each of these peptides derived from the Mexican,

15 Burmese and US strains of HEV were synthesized, coated on a solid phase and utilized in ELISA tests to determine the relative usefulness of these synthetic peptides.

As noted in Table 42, the amino acid identity between HEV US-1 and the Burmese, Mexican, and Pakistani strains of HEV range from about 87.5% to about 91.7% for the amino acids comprising the 3-2e epitopes within ORF 2, and from about 63.6 to about 72.7% for the

20 amino acids comprising the 4-2 epitopes within ORF 3. Without wishing to be bound by theory, given the degree of variability in the regions encoding for epitopes, it is likely that there may be strain specific antibody responses to these viruses.

TABLE 42 – (Similarity/Identify)

	3-2e Peptide			4-2 Peptide		
	Pak	Mex	US-1	Pak	Mex	US-1
Bur	100 / 97.9	91.7 / 91.7	93.7 / 91.7	100 / 100	72.7 / 72.7	72.7 / 72.7
Pak		91.7 / 91.7	93.7 / 91.7		72.7 / 72.7	72.7 / 72.7
Mex			89.6 / 87.5			63.6 / 63.6

B. Use of ELISA's in diagnosing acute HEV infection

It has been reported that most cases of acute HEV infection in man are accompanied by IgM class antibodies which bind to one or more HEV recombinant proteins or synthetic peptides. If a person does not have IgM class antibodies to HEV, the basis for diagnosis of acute HEV infection cannot be made on serology alone but may require, RT-PCR and/or other tests to verify HEV as the etiologic agent.

C. Generation of Synthetic Peptides

Peptides were prepared on a Rainin Symphony Multiple Peptide Synthesizer using standard Fmoc solid phase peptide synthesis on a 0.025 μ mole scale with (HBTU) coupling chemistry by in situ activation provided by N-methyl-morpholine, with 45 minute coupling times at each residue, and double coupling at predetermined residues. Standard cleavage of the resin provided the unprotected peptide, followed by ether precipitation and washing. The peptides synthesized are shown in Table 43.

TABLE 43

Peptide	Sequence	SEQ ID NO:
B 3-2e	TLDYPARAHTFDDFCPECRPLGLQGCAFAQSTVAELQRLKMKVGKTREL	SEQ ID NO:172
B 4-2	ANPPDHSAPLGVTRPSAPPLPHVVDLPQLGPRR	SEQ ID NO:171
M 3-2e	TFDYPGRAHTFDDFCPECRALGLQGCAFAQSTVAELQRLKVKVGKTREL	SEQ ID NO:170
M 4-2	ANQPGHLAPLGEIRPSAPPLPPVADLPQGLRR	SEQ ID NO:169
US 3-2e	TVDYPARAHTFDDFCPECRTLGVQGCAFAQSTIAEVQRLKMKVGKTREV	SEQ ID NO:174
US 4-2	DSRPAPSVPLGVTSPSAPPLPPVVDLPQLGLRC	SEQ ID NO:173

D. Analysis of Synthesized Peptides

The synthesized peptides were analyzed for their amino acid composition as follows. The crude peptides from the small scale syntheses (0.025 μ mole) were analyzed for their quality by C18 reverse phase high pressure liquid chromatography using an acetonitrile/water gradient with 0.1% (v/v) 2 trifluoroacetic acid (TFA) in each solvent. From the analytical chromatogram, the major peak from each synthesis was collected and the effluent analyzed by mass spectrometry (electrospray and/or laser desorption mass spectrometry. Purification of the peptides (small and/or large scale) was achieved using C18 reverse phase HPLC with an acetonitrile/water gradient with 0.1% TFA in each solvent. The major peak was collected, and lyophilized until use.

E. ELISA Test

The utility of the HEV US-1 epitopes was determined by coating 1/4 inch polystyrene beads with each peptide. Specifically, the peptides were solubilized in water or water plus glacial acetic acid and diluted to contain 10 μ g/mL in phosphate buffer (pH 7.4). A total of 60 polystyrene beads were added to a scintillation vial along with 14 mL of peptide solution (10 μ g/mL) and incubated at 56°C for two hours phosphate buffered saline (PBS). After incubation, the liquid was aspirated and replaced with a buffer containing 0.1% Triton-X100®. The beads were exposed to this solution for 60 minutes, the fluid aspirated and the beads washed twice with PBS buffer. The beads then were incubated with 5% bovine serum albumin solution for 60 minutes at 40°C. After incubation, the fluid was aspirated and the beads rinsed

with PBS. The resulting beads were soaked in PBS containing 5% sucrose for 30 minutes. The fluids then were aspirated and the beads air-dried.

In one study, one-quarter inch polystyrene beads were coated with various concentrations of the synthetic peptide (approximately 50 beads per lot) and evaluated in an ELISA test (described below) using serum from an anti-HEV seronegative human as a negative control and convalescent sera from an HEV-infected person as a positive control. The bead coating conditions providing the highest ratio of positive control signal to negative control signal were selected for scaling up the bead coating process. Two 1,000 bead lots were produced for both HEV US-1 ORF 2 and ORF 3 epitopes and then used as follows.

A sample of sera or plasma was diluted in specimen diluent and mixed with antigen-coated solid phase under conditions that permit an antibody in the sample to bind to the immobilized antigen. After washing, the resulting beads were mixed with horseradish peroxidase (HRPO)-labeled anti-human antibodies that bind to either tamarin or human antibodies bound to the solid phase. Specimens which produced signals above a cutoff value were considered reactive.

More specifically, the preferred ELISA format requires contacting the antigen-coated solid phase with serum pre-diluted with specimen diluent (buffered solution containing animal sera and non-ionic detergents). Specifically, 10 μ L of serum was diluted in 150 μ L of specimen diluent and vortexed. Then 10 μ L of this pre-diluted specimen was added to each well of an ELISA plate, followed by the addition of 200 μ L of specimen diluent and an antigen coated polystyrene beads. The ELISA plate then was incubated in a Dynamic Incubator (Abbott Laboratories) with constant agitation at room temperature for 1 hour. After the incubation, the fluids were aspirated, and the wells washed three times in distilled water (5 mL per wash). Next, 200 μ L of HRPO-labeled goat anti-human immunoglobulin diluted in a conjugate diluent (buffered solution containing animal sera and non-ionic detergents) was added to each well and the ELISA plate incubated for 1 hour, as indicated above. The wells then were washed three times in distilled water, the beads containing antigen and bound immunoglobulins removed from each well, and then placed in a test tube with 300 μ L of a solution of 0.1M citrate buffer (pH 5.5), 0.3% o-phenylenediamine-2 HCl and 0.02% hydrogen

peroxide. After 30 minutes at room temperature, the reaction was terminated by the addition of 1 N sulphuric acid. The resulting absorbance at 492 nm was the recorded. The intensity of the color produced was directly proportional to the amount of antibody present in the test sample. For each group of specimens, a preliminary cutoff value was set to separate specimens which
 5 presumably contained antibodies to the HEV epitope from those specimens which did not.

Panel 1: Testing of pre-screened panels

In order to demonstrate the utility of epitopes derived from the HEV US-1 strain, a panel of specimens was tested by an ELISA based on the HEV US-1 amino acid sequences (Table 44). These samples had been pre-screened for antibodies to HEV, using a combination of
 10 existing peptides and a licensed anti-HEV (Abbott Laboratories) as described above and in published reports (Dawson *et al.* (1993) *supra*; Paul *et al.* (1993) *supra*).

The first 10 members of the panel consisted of specimens obtained from US volunteer blood donors whose sera was negative for antibodies to HEV following analysis using a combination of peptides and recombinant proteins derived from Burmese and Mexican strains
 15 of HEV. All the specimens were non-reactive with ELISA's derived from HEV US-1. Five additional specimens were obtained from individuals suffering from acute hepatitis, and who were diagnosed with acute HEV infection because their sera was reactive for both IgG and IgM class antibodies to HEV recombinant antigens and synthetic peptides based on the Burmese and Mexican strains of HEV. Three of the five samples were from Egypt, one from India and one
 20 from Norway (a traveler). HEV RNA was detected by RT-PCR in all five of these individuals. These five members were tested for antibodies to the HEV US-1 isolate and both IgG and IgM class antibodies were detected in each of the cases (Table 44). Thus, these data support the use of synthetic peptides from the US-1 strain of HEV as having utility in diagnosing exposure to HEV and for diagnosing acute HEV infections.

TABLE 44

Test Specimens	Licensed anti HEV		US Isolate			
			IgG		IgM	
Tested	IgG	IgM	4-2	3-2e	4-2	3-2e
Neg. Control	0.061	0.084	0.031	0.041	0.071	0.109
Pos. Control	0.567	1.051	1.606	1.619	1.376	1.798
US Volunteer Donors						
TG 827	-	-	-	-	-	-
EG 549	-	-	-	-	-	-
EC 760	-	-	-	-	-	-
RF 762	-	-	-	-	-	-
RF 762	-	-	-	-	-	-
RG 730	-	-	-	-	-	-
NH 770	-	-	-	-	-	-
AS 705	-	-	-	-	-	-
BW 494	-	-	-	-	-	-
CD 648	-	-	-	-	-	-
Egypt						
7	+	+	+	+	+	+
9	+	+	+	+	+	+
12	+	+	+	-	+	+
India	+	+	+	+	+	+
543						
Norway						
M1	+	+	+	+	+	+

Panel 2: Detection of antibodies to HEV in biological source of HEV US-1 isolate

Serial bleeds were obtained from the patient described in Example 1, whose serum served as the biological source for the HEV US-1 strain. Based on serological data obtained for the Burmese and Mexican strains of HEV, this patient would have been misdiagnosed as HEV

negative because of the lack of detectable IgM class antibodies to HEV. However, both IgM class (Table 45) and IgG class (Table 46) antibodies to the HEV US-1 strain were detected on all four bleed dates (Tables 45 and 46. Had this patient's sera been analyzed for the presence of IgG and IgM class antibodies to the HEV US 3-2e and US 4-2 peptides, a positive diagnosis of acute HEV infection would have been made. This diagnosis is further supported by the observation that the individual had acute hepatitis and most importantly, had detectable HEV US-1 strain RNA in serum samples. These data indicate that synthetic peptides derived from the HEV US-1 strain may be useful in more accurately diagnosing acute infection due to HEV.

TABLE 45

Specimens	IgM: ORF 3 synthetic peptide 4-2			IgM: ORF 2 synthetic peptide 3-2e		
	ISOLATES			ISOLATES		
Tested	Burmese	Mexican	US-1	Burmese	Mexican	US-1
Negative Control	0.059	0.081	0.031	0.142	0.065	0.109
Positive Control	0.854	0.985	1.363	1.309	0.579	1.798
USP-1						
8 days post admission	-	-	+	-	-	+
9 days post admission	-	-	+	-	-	+
10 days post admission	-	-	+	-	-	+
37 days post admission	-	-	+	-	-	+

TABLE 46

Specimens	IgG: ORF 3 synthetic peptide 4-2			IgG: ORF 2 synthetic peptide 3-2e		
	ISOLATES			ISOLATES		
Tested	Burmese	Mexican	US-1	Burmese	Mexican	US-1
Negative Control	0.039	0.055	0.031	0.034	0.057	0.041
Positive Control	1.296	0.666	0.941	1.322	0.893	1.041
USP-1	-	-	+	-	-	+
8 days post admission	-	-	+	-	-	+
9 days post admission	-	-	+	-	-	+
10 days post admission	-	-	+	-	-	+
37 days post admission	-	-	+	-	-	+

Panel 3 - Other cases of potential acute HEV infection

A panel of sera from 50 patients diagnosed with acute hepatitis who were negative for IgM class antibodies to the Burmese and Mexican strains was assembled. Ten of 50 sera samples were positive for antibodies to the US strain of HEV (Tables 47 and 48). RT-PCR was performed on these samples, but none of the 10 were positive for HEV RNA. Thus, as demonstrated in this example, when patient sera is analyzed for the presence of antibodies to HEV US-1, occult viral hepatitis may be diagnosed as acute HEV infection.

TABLE 47

Specimens	IgM: ORF 3 synthetic peptide 4-2			IgM: ORF 2 synthetic peptide 3-2e		
	ISOLATES			ISOLATES		
Tested	Burmese	Mexican	US-1	Burmese	Mexican	US-1
Negative Control	0.059	0.081	0.031	0.142	0.065	0.109
Positive Control	0.854	0.985	1.363	1.309	0.579	1.798
US	-	-	-	-	-	+
Acute non A-E	-	-	-	-	-	+
SH 755	-	-	-	-	-	+
DT 314	-	-	-	-	-	+
EH 673	-	-	-	-	-	+
SG560	-	-	-	-	-	+
SR681	-	-	-	-	-	-
N11C10	-	-	+	-	-	+
35	-	-	+	-	-	+
52	-	-	-	-	-	+
161	-	-	-	-	-	+
175						

TABLE 48

Specimens	IgG: ORF 3 synthetic peptide 4-2			IgG: ORF 2 synthetic peptide 3-2e		
	ISOLATES			ISOLATES		
Tested	Burmese	Mexican	US-1	Burmese	Mexican	US-1
Negative Control	0.039	0.055	0.031	0.034	0.057	0.041
Positive Control	1.296	0.666	0.941	1.322	0.893	1.041
US	-	-	-	-	-	-
Acute non A-E	-	-	-	-	-	-
SH 755	-	-	-	-	-	-
DT 314	-	-	-	-	-	-
EH 673	-	-	-	-	-	-
SG560	-	-	-	-	-	-
SR681	-	-	-	-	-	+
N11C10	-	-	-	-	-	-
35	-	-	-	-	-	+
52	-	-	-	-	-	-
161	-	-	-	-	-	-
175						

Example 9 - Animal Transmission Studies

Cynomolgus macaques (*Macaca fascicularis*) were obtained through the Southwest Foundation for Biomedical Research (SFBR) in San Antonio, Texas. The animals were maintained and monitored in accordance with guidelines established by SFBR to ensure

- 5 humane care and the ethical use of primates. Sera were obtained twice weekly for at least four weeks prior to inoculation in order to establish the baseline levels for serum ALT. Cut-off (CO) values were determined based on the mean of the baseline plus 3.75 times the standard deviation. Two macaques were inoculated intravenously with 0.4-0.625 mL of HEV positive USP-1 serum and one macaque was inoculated with 2.0 mL of HEV positive USP-2 serum.
- 10 Serum and fecal samples were collected twice weekly for up to 16 weeks post-inoculation (PI). Sera were tested for changes in ALT and values greater than the CO were considered positive and suggestive of liver damage. Sera samples were tested for antibodies to HEV as described

hereinabove in Example 8 (Table 49, Figure 7). Sera and fecal samples were tested for HEV RNA by RT-PCR. 25-100 μ L of macaque sera was extracted using the QIAamp Viral RNA Kit (Qiagen). 10% fecal suspension were extracted as described in Example 1. RT PCR was performed as described below in Example 12 (Figure 7).

5 Although intravenous inoculation of 0.4-0.625 mL of USP-1 sera into two cynomolgus macaques failed to produce infection (data not shown), inoculation of 2.0 mL of sera from patient US-2 resulted in viremia and elevations of liver enzyme levels in the serum (Figure 7). HEV RNA was first detected in fecal material on day 15 PI and remained positive through 64 days PI. Serum specimens collected between days 28-56 PI were HEV RNA positive. Elevated
10 ALT values were noted on days 15, 44-58, 72 and 93 PI, with the peak ALT value (116 IU/L) on day 51 PI.

 Six ELSIAs based on the Burmese, Mexican and US sequences for the 4-2 and 302e peptides were utilized to assess antibody response. Measurable response was found only to the US 3-2e peptide assay (Table 49) with no noted crossreactivity to the Burmese or Mexican
15 peptides. IgM class antibody directed against HEV was detectable between 28 and 58 days PI. This was followed by a strong anti-HEV-IgG response at day 44 PI.

TABLE 49

Date	DPI	ALT	AST	GGT	IgG S/N
06/04/97	-82	35	37	102	1.4
06/06/97	-80	39	32	90	
06/11/97	-75	38	36	100	
06/13/97	-73	36	46	86	
06/18/97	-68	45	30	85	
06/20/97	-66	43	37	87	
06/25/97	-61	37	30	92	
06/27/97	-59	42	36	87	
08/25/97	0	41	36	107	1
08/27/97	2				
09/02/97	8	34	34	102	
09/04/97	10	34	31	91	
09/09/97	15	58	42	108	0.8
09/10/97	16	44	45	93	
09/15/97	21	35	32	86	
09/17/97	23	49	71	88	
09/22/97	28	39	33	86	1.2
09/24/97	30	40	37	90	
09/29/97	35	41	40	80	
10/01/97	37	48	58	90	1.1
10/03/97	39				
10/06/97	42	45	33	89	
10/08/97	44	58	38	94	6.2
10/15/97	51	116	62	89	11.9
10/20/97	56	87	38	83	33.6
10/22/97	58	76	43	85	29.9
10/28/97	64	45	42	88	17.2
10/29/97	65	46	34	88	
11/03/97	70	39	54	85	
11/05/97	72	54	47	88	13.3
11/10/97	77	47	33	93	
11/12/97	79	50	38	93	12.4
11/17/97	84	46	31	91	10.4
11/19/97	86	52	41	88	
11/26/97	93	67	104	109	7.2
12/03/97	100	36	36	108	
12/09/97	106	38	34	115	
12/10/97	107	36	29	103	2.1

Example 10: Recombinant Protein ELISAs

A. Recombinant Constructs

E. coli derived recombinant proteins encoded by HEV-US sequence from the ORF 2 and ORF 3 regions of the HEV-US genome were expressed as fusion proteins with CMP-KDO synthetase (CKS), designated as pJOorf3-29 (SEQ ID NO:191); cksorf2m-2 (SEQ ID NO:192); and CKSORF32M-3 (SEQ ID NO:193), or as non-fusion proteins, designated as plorf3-12 (SEQ ID NO:194); plorf2-2.6 (SEQ ID NO:195); and PLORF-32M-14-5 (SEQ ID NO:196). The cloning vector pJO201, as described in U.S. Patent No. 5,124,255, was used in the construction of the recombinant fusion proteins. This vector was digested with the restriction endonucleases *Eco RI* and *Bam HI* to allow cloning of HEV-US sequences in frame with CKS. The lambda pL expression vector pKRR826 was utilized in the construction of recombinant non-fusion proteins. This vector was digested with the restriction endonucleases *Eco RI* and *Bam HI* to allow for cloning of HEV-US sequences immediately down stream of the ribosome binding site. Since the vector system contains strong lambda promoter, induction of heterologous protein synthesis is accomplished by shift in the temperature from 30°C to 42°C which inactivates the temperature sensitive repressor protein. The constructs were cloned and transformed into *E. coli* K12 strain HS36 cells for the expression of these HEV proteins.

HEV-US sequences were amplified from nucleic acids extracted from HEV US-2 human serum or macaque 13906 fecal material and reverse transcribed as described above in Example 5. The ORF 2 sequence, encompassing the carboxyl half of ORF 2 (*i.e.*, encoding amino acid residue numbers 334-660 of SEQ ID NO:167), was generated using a sense primer, SEQ ID NO:208, which contained an *Eco RI* restriction site as well as an ATG start codon and an antisense primer, SEQ ID NO:198, which contained a unique peptide sequence termed FLAG (Eastman Kodak), two consecutive TAA termination codons, and a *Bam HI* restriction site. A 50 µl PCR reaction was set up using LA TAQ (Takara) reagents as recommended by the manufacturer. Cycling conditions involved 40 cycles of 94°C for 20 seconds, 55°C for 30 seconds, 72°C for 2 minute. Amplifications were preceded by 1 minute at 94°C and followed by 10 minutes at 72°C. Products were digested with *Eco RI* and *Bam HI* and ligated into the

desired vector. The nucleotide sequence of the CKS fusion clone, between the restriction sites, is set forth in SEQ ID NO:192, the translation of which is set forth in SEQ ID NO:199. The nucleotide sequence of the non-fusion clone, between restriction sites, is set forth in SEQ ID NO:195, the translation of which is set forth in SEQ ID NO:200. The ORF 3 sequences, encompassing the entire ORF 3 (amino acids 1-122), was generated using a sense primer, SEQ ID NO:201, which contained an *Eco RI* restriction site as well as an ATG start codon and an antisense primer, SEQ ID NO:202, which contained a unique peptide sequence termed FLAG, two consecutive TAA termination codons, and a *Bam HI* restriction site. A 50 μ L PCR reaction was set up using Qiagen reagents as described in Example 5. Cycling conditions comprised 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 1 minute. Amplifications were preceded by incubation for 1 minute at 94°C, followed by 10 minutes at 72°C. The resulting products were digested with *Eco RI* and *Bam HI* and ligated into the desired vector. The nucleotide sequence of the CKS fusion clone, between the restriction sites, is set forth in SEQ ID NO:191, the translation of which is set forth in SEQ ID NO:203. The nucleotide sequence of the clone representing the non-fusion construct, between the restriction sites, is set forth in SEQ ID NO:195, the translation of which is set forth in SEQ ID NO:204.

Additionally, a chimeric construct encompassing the full length ORF 3 (amino acids 1-123) and the carboxyl half of ORF 2 (amino acids 334-660) was generated. Approximately 100 ng of the plasmids containing SEQ ID NO:191 and SEQ ID NO:192 were utilized as template in 100 μ L PCR reactions. PCR buffers and enzymes were from the LA TAQ kit (Takara), and used in accordance with the manufacturer's instructions. ORF 3 was amplified with primers set forth in SEQ ID NOS:201 and 205. The antisense primer of SEQ ID NO:205 eliminates the FLAG sequences and stop codons from the carboxyl end of SEQ ID NO:191 and contains the sequence identical to SEQ ID NO:192 which will eliminate the ATG start codon. ORF 2 was amplified with primers of SEQ ID NOS:208 and 198. Cycling conditions were as described above using LA TAQ. The resulting products were fractionated on a 1.2% agarose gel and excised. DNA was isolated from the gel slices using GeneClean II as described by the manufacturer (Bio101). Products were eluted off the glass beads into 15 μ L H₂O. Approximately equal molar ratios of each product (10 μ L of ORF 3 product and 1 μ L of ORF 2 product) were mixed in a 25 μ L end fill reaction using 1x PCR buffer, 0.5 μ L dNTPs, and 0.25

5 μ L LA TAQ (Takara). This reaction was cycled as follows: 94°C for 1 minute, 10 cycles of 94°C for 20 seconds, 55°C for 30 seconds, and 72°C for 1.5 minutes, followed by 72°C for 10 minutes. 5 μ L of this reaction was placed into a 100 μ L amplification reaction utilizing LA TAQ kit (Takara) and primers of SEQ ID NOS:201 and 198. Cycling conditions were 94°C for 1 minute followed by 35 cycles of 90°C for 20 seconds, 55°C for 30 seconds, and 72°C for 1.5 minutes. This was followed by 10 minutes at 72°C and a 4°C soak. Products of the appropriate size were digested with restriction enzymes *Eco RI* and *Bam HI*. This product was ligated into pJO201 and clones with the appropriate sequence identified (SEQ ID NO:193, the translation of which is set forth in SEQ ID NO:206). The resulting product was ligated into pKRR826 and clones with the appropriate sequence (SEQ ID NO:196, the translation of which is set forth in SEQ ID NO:207) identified.

B. Protein expression and purification

15 The CKS constructs were expressed in two 500 mL cultures (4 hour induction), as described in U. S. Patent No. 5,312,737. P_L constructs were expressed as described above. Frozen cell pellets of the induced *E.coli* cultures were used as the starting material for the purification of protein. Cells were lysed in buffer containing lysozyme, DNase and proteinase inhibitors. Soluble protein was separated from insoluble (inclusion body) protein by centrifugation at 11,000 x g. The solubility of the recombinant protein was estimated via sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) and Western blotting
 20 using a FLAG® M2 antibody.

25 Soluble recombinant protein was purified by affinity chromatography using FLAG® M2 antibody affinity gel after exchange into suitable buffer (Surowy *et al.* (1997) Journal of General Virology, 78:1851-1859). If necessary, additional purification was performed via Sephacryl® S-200 gel filtration chromatography, in which the sample and chromatography buffers contained 10 mM β -mercaptoethanol. Purified protein was quantitated by measurement of absorbance at 280 nm. An assumed extinction coefficient of 1 was used to convert absorbance to mg of protein. Protein purity was determined by scanning densitometry

(Molecular Dynamics) of protein fractionated by SDS PAGE, using standards of pre-determined purity.

C. ELISA

In order to determine potential utility of the recombinant HEV US constructs, solid phase ELISA's were developed and evaluated. All recombinant HEV US proteins were coated onto solid phase as described below. Briefly, 1/4" polystyrene beads were coated with varying amounts of (pJOORF3-29) which ranged in concentration from 0.5 to 10 µg/mL diluted in 100 mM sodium phosphate buffer, pH 7.6. Sixty beads per concentration condition were coated in approximately 14 mL of buffer and rotated end-over-end at 40° C for 2 hours. The coating solution was aspirated and the remainder of the coating procedure was performed as described above in Example 8, section E, paragraph 1.

An ELISA was developed using the pJOorf3-29 coated beads. Briefly, sera or plasma was diluted 1:16 in Specimen Diluent (SpD) as described above. A 10 µL aliquot of this pre-dilution then was added into the well of a reaction tray, followed by the addition of 200 µL of SpD. One coated bead was added per well and incubated for 1 hour at 37°C in dynamic mode using a Dynamic Incubator (Abbott Laboratories). After incubation, the fluid was aspirated and each bead washed 3 times with deionized water (5 mL per wash). The beads then were incubated with 200 µL HRPO-labeled goat anti-human IgG or IgM conjugate, diluted in conjugate diluent (described above) and incubated for 30 minutes at 37°C. The conjugate then was aspirated and the beads washed as above. Color development and absorbance readings were performed as described in Example 8, section E.

To validate the immunoreactivity of this construct, serial bleed specimens from Macaque #13903 experimentally infected with HEV US-2 (described in Example 9) were tested for IgM and IgG antibody to pJOorf3-29. As shown in Figure 1, IgM antibody was detected at day 51 post-infection (PI) and continued to be elevated through day 72 and corresponded to the peak elevations in ALT values. IgG antibody to pJOorf3-29 was first detected on day 56 PI and remained positive through day 107 (Table 50).

A second construct, plorf3-12, representing HEV US ORF 3 but lacking the CKS fusion partner was also evaluated in an ELISA format identical to that described above. IgG antibody to plorf3-12 was evaluated on several serial bleeds from the same experimentally infected macaque. IgG antibody to plorf3-12 was detected on day 58 PI and remained positive through day 107 (Table 50).

TABLE 50

Sample	pJOorf3-29		plorf3-12	
	Mean OD	S/N	Mean OD	S/N
SpD			0.01	
"pre-bleed"	0.02		0.01	
Post-inoculation bleeds - Days Post-inoculation (DPI)				
DPI				
44	0.02	0.96	0.02	1.07
51	0.05	2.35	0.03	2.25
56	0.24	10.35	0.05	3.43
58	0.44	19	0.16	11.57
63	1.14	49.57	0.32	22.82
65		NT	0.53	37.54
70		NT	1.19	85.04
72	2.22	96.52	0.92	65.71
98	0.89	38.87	0.39	27.86
107	0.49	21.43	0.27	19.36
NT: not tested				

Due to the high percent homology between Swine HEV and the US-2 isolate, the pJOorf3-29 ELISA also was used to measure the prevalence of both immunoreactive IgG and IgM in sera isolated from U.S. swine herds (Table 51). The assay was performed as described above with the exception of substituting HRPO-conjugated labeled anti-swine immunoglobulin (either IgG or IgM) for the anti-human conjugate.

TABLE 51

Prevalence of Antibody to HEV orf3 in U. S. Swine (pJOorf3-29)					
Swine Source State	IgG Reactive No./Total (%)	No. IgG Confirmed by Blocking or Blot (%)	IgM Only Reactive No./Total (%)	No. IgM Only Confirmed by Blot (%)	Total Exposure Confirmed Only
New Jersey	9/14 (64)	9 (100)	0/14		64%
Texas	25/50 (50)	20 (80)	0/50		40%
Iowa	7/64 (11)	1 (14)	0/64		2%
Oregon	7/36 (19)	5 (71)	1/36 (3)	1/1 (100)	14%
Total	48/164 (29)	35 (73)	1/164 (0.6)	1/1 (100)	36/164 (22%)
NOTE: A total of 4 pigs (all Texas herd) had IgM in addition to IgG.					

In order to confirm reactive specimens, a blocking assay was developed. Briefly, a 10 μ L aliquot of the 1:16 specimen pre-dilution was added to duplicate wells of a reaction tray; one well to be used for the standard assay and one well to be used for the blocking assay. The ELISA for the standard assay was performed as described above with the exception that there was a 30 minute room temperature pre-incubation step prior to addition of the pJOorf3-29 antigen coated bead. For the blocking assay, pJOorf3-29 was added to the SpD (blocking reagent) at a 10-fold molar excess to that on the solid phase. 200 μ L of blocking reagent was added per reaction and a 30 minutes room temperature pre-incubation was performed prior to addition of the pJOorf3-29 antigen coated bead. The rest of the assay was performed as described above for the swine assay, except that the HRPO-conjugated anti-swine conjugate (IgG) was used in place of the anti-human conjugate.

The % blocking was determined using the equation:

$$[(A_{492 \text{ nm}} \text{ standard assay} - A_{492 \text{ nm}} \text{ blocking assay}) / A_{492 \text{ nm}} \text{ standard assay}] \times 100$$

Specimens that showed blocking rates of 50% or greater were considered to be reactive for IgG antibody to HEV pJOorf3-29. Representative IgG positive and IgG negative swine samples and their blocking results are shown in Table 52.

Table 52 - Blocking Assay With pJOorf3-29 and PL-12 at 10-fold molar excess

Standard Assay			Blocking Assay w/ pJOorf3-29 at 10-fold molar excess			BLOCKING RESULT
SAMPLE	OD	MEAN OD	OD	MEAN OD	% BLOCKING	
NC	0.02	0.02	0.02	0.02	50.4%	+
	0.02		0.03			
	1.09		0.56			
PC	1.01	1.05	0.48	0.52		
Oregon Swine Panel Positives						
1	NJ5	0.65	0.15		76.5%	+
2	NJ12	1.78	0.46		74.0%	+
3	NJ21	0.48	0.16		66.7%	+
4	NJ23	0.52	0.09		81.9%	+
5	T5	2	0.81		59.5%	+
6	T9	0.52	0.18		64.3%	+
7	T32	2	0.9		54.9%	+
8	T33	0.3	0.13		57.8%	+
9	T48	0.53	0.14		73.7%	+
10	T49	0.33	0.09		73.3%	+
Oregon Swine Panel Negatives						
11	T43	0.08	0.07		13.3%	-
12	T46	0.12	0.08		29.1%	-
13	I-23	0.12	0.08		32.2%	-
14	I-24	0.07	0.06		13.2%	-
15	I-27	0.1	0.08		12.6%	-
16	I-28	0.15	0.12		20.4%	-
17	I-33	0.15	0.12		19.9%	-
18	I-39	0.23	0.14		37.4%	-
19	I-61	0.19	0.14		25.9%	-
20	O-4	0.15	0.12		22.7%	-

In addition to the blocking assay, western blots were run on a subset of swine specimens. Briefly, 50 µg of HEV pJOorf3-29 and 50 µg of "CKS only" proteins were fractionated by SDS-PAGE and the fractionated proteins transferred to nitrocellulose. 3mm strips of the nitrocellulose were cut and incubated overnight at room temperature on an orbital rotator with primary antibody at a 1:100 dilution in protein based buffer containing 10% *E. coli* lysate. On the following day, strips were washed three times with 0.3% Tween/TBS (TBST), followed by the addition of HRPO-conjugated anti-swine IgG conjugate diluted to 0.5 µg/mL in TBST. Strips were incubated with rotation for 4 hours at room temperature. Blots then were washed three times in TBST, followed by 2 washes in TBS. Blots were developed using 4-chloro-1-naphthol as a substrate. The reaction was stopped by the addition of water and band intensities recorded. Specimens were determined to have specific reactivity to HEV if they showed a band at the correct molecular weight for pJOorf3-29 (approx. 40 kD) and had no reactivity in the region where "CKS only" bands (approx. 29 kD). Results for 20 swine sera run on the pJOorf3-29 western blot are shown in Table 53. No swine sera showed non-specific reactivity with the "CKS-only" band.

TABLE 53

Swine ID Number	BAND INTENSITY	
	pJOorf3-29	CKS only
NJ4	+	-
NJ7	+	-
NJ14	+++	-
NJ18	+	-
NJ25	++++	-
T6	++++	-
T10	++++	-
T14	-	-
T15	+	-
T18	++	-
T28	+++	-
T29	-	-
T30	+	-
T34	-	-
T36	++++	-
T37	-	-
T43	-	-
T44	++++	-
T45	++++	-
T46	-	-

These data suggest that HEV US recombinant proteins are useful in diagnosing exposure to HEV.

Example 11 - Consensus Primers

Consensus oligonucleotide primers for HEV ORF 1 ORF 2 and ORF 3 were designed based on conserved regions between the full length sequences of isolates from Asia, Mexico, and the US (Figure 9). The ORF 1 primers are positioned within the methyltransferase region at nucleotides 56-79 and 473-451 of the Burmese isolate (GenBank accession number M73218), and amplify a product 418 nucleotides in length. The ORF 1 primers include:

HEVConsORF 1-s1; CTGGCATYACTACTGCGYATTGAGC (SEQ ID NO:147); and
HEVConsORF 1-a1; CCATCRARRCAGTAAGTGCGGTC (SEQ ID NO:148).

The ORF 2 primers, at positions 6298-6321 and 6494-6470 of the Burmese isolate, produce a product 197 nucleotides in length. The ORF 2 primers include:

5 HEVConsORF 2-s1; GACAGAATTRATTTCGTCGGCTGG (SEQ ID NO:150); and
HEVConsORF 2-a1; CTTGTTCRTGYTGGTTRTCATAATC (SEQ ID NO:126).

For a second round of amplification, internal primers can be used to produce products 287 and 145 nucleotides in length for ORF 1 and ORF 2, respectively. The ORF 1 primers
10 include:

HEVConsORF 1-s2; CTGCCYTKGCGAATGCTGTGG (SEQ ID NO:177); and
HEVConsORF 1-a2; GGCAGWRTACCARCGCTGAACATC (SEQ ID NO:178).

The ORF 2 primers include:

15 HEVConsORF 2-s2; GTYGTCTCRGCCAATGGCGAGC (SEQ ID NO:152); and
HEVConsORF 2-a2; GTTCRTGYTGGTTRTCATAATCCTG (SEQ ID NO:128).

PCR reactions contained 2 mM MgCl₂ and 0.5 μM of each oligonucleotide primer as per the manufacturer's instructions (Perkin-Elmer) and amplified using Touch-down PCR as described in Example 5. Amplified products were separated on a 1.5% agarose gel and analyzed for the presence of PCR products of the appropriate size. The primers were used to
20 detect the presence of virus in serum and feces containing HEV US-2 as described above in Example 8 and Figure 7. In addition, these primers were found to be reactive with a number of different variants of HEV that included Burmese-like strains 6A, 7A, 9A and 12 A as well as two distinct isolates from Greece (see Example 13 below) as well as a unique isolate from Italy and the two isolates from the US (see Example 13 below). In addition, these primers have been
25 used to identify an isolate from a patient with a clinical diagnosis of acute sporadic hepatitis from the Liaoning province of China (S15). The results are presented in Table 54 below.

TABLE 54

Sample	ORF 1 -PCR1	ORF 1 -PCR 2	ORF 2 - PCR1	ORF 2 -PCR2
6A	neg	pos	pos	Pos
7A	neg	pos	neg	Pos
9A	neg	neg	neg	Pos
12A	pos	pos	neg	Neg
G1	pos	pos	pos	Pos
G2	pos	pos	pos	Pos
Itl	pos	pos	pos	Pos
S15	nd	pos	nd	Pos
US-2	pos	pos	pos	Pos

Example 12 - Detection of HEV RNA in Primary Human Fetal Kidney Cells

Frozen cell pellets containing 10×10^6 cells were thawed and resuspended in 1.0 mL Dulbecco's phosphate buffered saline. RNA was extracted from 20 μ L (2×10^5 cells) of the cell pellet using the Ultraspec Isolation System as described in Example 1. cDNA synthesis was performed on the above extracted nucleic acid (RNA) and primed with random hexamers. PCR then was performed on the above cDNA using degenerate primers from the ORF-1 and ORF-2 regions of the viral genome at a final concentration of 0.5 μ M as described in Example 11.

To monitor the performance of the above assay, a positive control utilizing primary human kidney cells and HEV US-2 positive serum was included in the experimental design. Two positive control sets were prepared by spiking 2×10^5 HEV negative primary human kidney cells with 2.5 μ L and 25 μ L of a documented HEV US-2 positive serum specimen. The positive control serum also was tested without the addition of the human kidney cells.

Nineteen primary human kidney cell pellet lots were tested using the above assay method utilizing the 2 degenerate primer sets from ORF 1 and ORF 2. The results are summarized in Table 55 below. None of the cell pellet lots tested gave positive results as seen in the positive controls.

TABLE 55

CELL LINES	PCR RESULTS
1757	-
1851	-
1690	-
1853	-
1906	-
1935	-
1838	-
1955	-
1893	-
1895	-
1699	-
1877	-
1942	-
1844	-
1840	-
1875	-
1921	-
1946	-
1846	-
cells + 25 µL serum	+
cells + 2.5 µL serum	+
25 µL serum	+

Example 13: Identification and Extension of Additional US-type Isolates

A. Identification of isolate from Italy, referred to as It1

RNA was extracted from 25 to 50 µL of serum using the QIAamp Viral RNA kit (Qiagen) as described by the manufacturer except that 25 to 50µL of serum was diluted to 100µL with PBS and the final elution was performed with 100 µL of RNase-free water. RT reactions were random primed. PCR utilized the HEV US-1 primer as described hereinabove in Example 5. A 294 bp product was generated after amplification with primers SEQ ID NO:94

and SEQ ID NO:96. The product was cloned and sequenced as described in Example 3 and is shown in SEQ ID NO:179.

Extension of the Itl isolate genome was performed as follows. RNA was extracted from 25 to 50 μ L of serum as described hereinabove in Example 5. RT reactions were random primed. PCR utilized the HEV CONSENSUS primers described above in Example 11 using touchdown PCR, as described hereinabove in Example 3. Primers shown in SEQ ID NOS:147 and 148 were used to generate a product having the sequence set forth in SEQ ID NO:180 (reaction z2, 418 bp). Primers as shown in SEQ ID NOS:150 and 126 were used to generate a product having the sequence set forth in SEQ ID NO:181 (reaction z3, 197 bp). In the presence of 1x PCR Buffer and 20% Q Solution (Qiagen), primers as shown in SEQ ID NOS:182 and 183 were used to generate a product having a sequence set forth in SEQ ID NO:184 (reaction z4, 234 bp). The 3' end of the genome was isolated by 3' RACE as described above in Example 3 using primers shown in SEQ ID NOS:150 and 85 in PCR1, and primers shown in SEQ ID NOS:152 and 85 in PCR2, to produce a product having the sequence shown in SEQ ID NO:185 (reaction z5, 890 bp). Products were cloned and sequenced as described in Example 3 and consensus sequences generated. These regions are shown in Figure 8 and are set forth in SEQ ID NOS:180, 184 and 186. The amino acid translations of these regions are represented by the amino acid sequences set forth in SEQ ID NOS:187, 188; 189; 190; and 197.

B. Identification of two isolates from Greece, referred to as G1 and G2

Two patients with acute hepatitis who had no history of travel to endemic areas had been analyzed with primers based on the Burmese isolate (Psichogiou M.A., *et al.*, (1995) "Hepatitis E virus (HEV) infection in a cohort of patients with acute non-A, non-B hepatitis," *Journal of Hepatology*, 23, 668-673). Only patient G2 was found to be PCR positive. RNA was isolated as described hereinabove in Example 12 and PCR performed with the consensus primers described above in Example 11. The ORF 1 and ORF 2 primer sets generated products of the expected size from both patients. The products were cloned and sequenced as described above in Example 3. The products generated using the ORF 1 and ORF 2 consensus primers from patient G1 are shown in SEQ ID NOS:209 and 211, respectively. The products generated

using the ORF 1 and ORF 2 consensus primers from patient G2 are shown in SEQ ID NOS:213 and 215, respectively. The identification of G1 as being PCR positive demonstrates the utility of the consensus primers over Burmese base strain specific primers.

Additional sequence from G1 and G2 was also obtained using primers SEQ ID NO:16, SEQ ID No:17, and SEQ ID NO:18 as for the generation of SEQ ID NO:19 as described above in Example 3 except that random primed cDNA was used for PCR and amplification involved 10 cycles of 94°C for 20 seconds, 60°C for 30 seconds, and 72°C for 1 minute, followed by 10 cycles of 94°C for 20 seconds, 55°C for 30 seconds, and 72°C for 1 minute followed by 30 cycles of 94°C for 20 seconds, 50°C for 30 seconds (-0.3°C/cycle), and 72°C for 1 minute. This was followed by an extension cycle of 72°C for 7 minutes. The product generated from patient G1 is shown in SEQ ID NO:217. The product generated from patient G2 is shown in SEQ ID NO:220.

Alignments of the nucleotide sequences of the US, Chinese, Greek, Italian, Mexican and Burmese-like isolates, were performed to determine the relationship of these isolates to each other. The divergence of the Italian isolate is supported by the comparisons of the product from the ORF 1 region of the genome which has a percent nucleic acid identity of 77.6 %, 78.4 %, and 84.6 % with the prototype isolates from Burma (B1), Mexico (M1) and the US (US-1), respectively (Table 36). The divergence of the Italian isolate also is supported by the comparisons of the product from the ORF 2 region of the genome which had a percent nucleic acid identity of 83.3 %, 79.7 %, and 87.8 % with the prototype isolates from Burma, Mexico and the US, respectively (Table 37). The nucleotide identities between the prototype isolates from Burma, Mexico and the US, range between 75.5 % to 82.4 % over these two regions. Over these same regions, the isolates that comprise the Burmese-like group have much higher identities of 91.2% or greater.

Comparisons of the ORF 1 and ORF 2 amplified sequences indicate that the isolates from the two patients from Greece are quite distinct from each other, exhibiting 84.4 % and 87.2 % nucleotide sequence identity over these regions of ORF 1 and ORF 2, respectively. At the nucleotide level, the percent identities between the Greek, Italian and US isolates range from 81.9% to 86.8% for the ORF 1 product (Table 36) and 82.4% to 87.8% for the ORF 2 product

(Table 37). These values are lower than the lowest percent nucleotide identities between any Burmese-like isolates, which are greater than 91.2% for both ORF 1 and ORF 2. Comparisons of the amino acid identities derived from the ORF 1 fragment between the US, Italian or Greek isolates and the Burmese or Mexican isolates range from 87.8% to 93.5 % (Table 36). These values are equal to or less than the differences between the Burmese and Mexican isolates (93.5% to 95.1 %) (Table 36), indicating that the isolates from non-endemic regions are distinct from the isolates originating from endemic regions.

The relative evolutionary distances between the viral sequences analyzed are readily apparent upon inspection of the unrooted phylogenetic trees generated from the pairwise distances, where the branch lengths are proportional to the relative genetic relationships between the isolates. The phylogenetic trees based on alignments of either ORF 1 (Fig. 10) or ORF 2 (Fig. 11) sequences are quite similar in overall topology. The Burmese-like isolates and the Mexican isolate represent major branches at one end of the tree. The human US isolates form a distinct group distal to the Mexican and Burmese isolates. The swine HEV-like sequence from ORF 2 is closely related to the US human isolates. The three European isolates form three additional distinct branches with the Italian isolate being most closely related to the US isolates.

Example 14: Identification Additional US-type Isolates from Austria and Argentina

RNA was isolated from serum from three patients with acute hepatitis who had no history of travel to areas considered endemic for HEV as described hereinabove in Example 12 and PCR performed with the consensus primers described above in Example 11. One patient was from Austria, Au1, (Worm, *et al.*, (1998) "Sporadic hepatitis E in Austria," New England Journal of Medicine, 339, 1554-1555) while the other two patients were from Argentina. The ORF 1 and ORF 2 primer sets generated products of the expected size from all patients. The products were cloned and sequenced as described above in Example 3. The products generated using the ORF 1 and ORF 2 consensus primers from patient Au1 are shown in SEQ ID NOS:243 and 245, respectively. The products generated using the ORF 1 and ORF 2 consensus primers from patient Ar1 are shown in SEQ ID NOS:247 and 249, respectively. The

products generated using the ORF 1 and ORF 2 consensus primers from patient Ar2 are shown in SEQ ID NOS:251 and 253, respectively. PCR products were obtained after both the first round of ORF1 PCR with the a1 and s1 primers as well as the second round of nested ORF1 PCR with the a2 and s2 primers for Au1, Ar1 and Ar2. PCR products were obtained after both the first round of ORF2 PCR with the a1 and s1 primers as well as the second round of nested ORF2 PCR with the a2 and s2 primers for Au1 and Ar2. Product from Ar1 was detected only after the second round of nested ORF2 PCR with the a2 and s2 primers.

Alignments of the nucleotide sequences of the US, Chinese, Greek, Italian, Austrian, Argentine, Mexican and Burmese-like isolates, were performed to determine the relationship of these isolates to each other as described in Example 6. The divergence of the Austrian isolate, Au1, is supported by the comparisons of the product from the ORF 1 region of the genome which has a percent nucleic acid identity of 77.1 %, 78.2 %, and 87.9 % with prototype isolates from Burma (B1), Mexico (M1) and the US (US-1), respectively (Table 56). The divergence of the Austrian isolate also is supported by the comparisons of the product from the ORF 2 region of the genome which had a percent nucleic acid identity of 85.1 %, 79.1 %, and 83.1 % with the prototype isolates from Burma (B1), Mexico (M1) and the US (US-1), respectively (Table 57). The divergence of the Argentine isolate, Ar2, is supported by the comparisons of the product from the ORF 1 region of the genome which has a percent nucleic acid identity of 76.0 %, 76.0 %, and 84.9 % with the prototype isolates from Burma (B1), Mexico (M1) and the US (US-1), respectively (Table 56). The divergence of the Ar2 isolate also is supported by the comparisons of the product from the ORF 2 region of the genome which had a percent nucleic acid identity of 85.8 %, 82.4 %, and 85.8 % with the prototype isolates from Burma (B1), Mexico (M1) and the US (US-1), respectively (Table 57). The divergence of the Argentine isolate, Ar1, is supported by the comparisons of the product from the ORF 1 region of the genome which has a percent nucleic acid identity of 76.6 %, 77.6 %, and 85.7 % with the prototype isolates from Burma (B1), Mexico (M1) and the US (US-1), respectively (Table 56). The nucleotide identities between the prototype isolates from Burma (B1), Mexico (M1) and the US (US-1), range between 75.5 % to 82.4 % over these two regions. Over these same regions, the isolates that comprise the Burmese-like group have much higher identities of 91.2% or greater. Although only a nested ORF2 PCR product was obtained from the Argentine isolate, Ar1, the

divergence of the Ar2 isolate also is supported by the comparisons of this smaller product from the ORF 2 region of the genome which had a percent nucleic acid identity of 80.6 % with the prototype isolates from Burma (B1), Mexico (M1) and the US (US-1) (Table 57).

At the nucleotide level, the percent identities between the Austrian, Argentine, Greek,
5 Italian and US isolates (excluding the identity between US-1 and US-2) range from 80.6% to 89.8% for the ORF 1 product (Table 56). At the nucleotide level, the percent identities between the Austrian, Argentine, Greek, Italian and US isolates (excluding the identity between US-1 and US-2 and Ar-1 and Ar-2) range from 80.6% to 89.2% for the ORF 2 product (Table 57).
10 These values are lower than the lowest percent nucleotide identities between any Burmese-like isolates, which are 91.2% or greater for ORF 1 and ORF 2.

Nucleotide and deduced amino acid identity between isolates of HEV over 371 base (123 amino acid) ORF 1 fragment

Nucleotide Identity																		
Ar1	88.4	89.8	84.4	81.9	85.4	85.7	85.2	82.5	76.6	76.6	79.0	78.2	79.2	77.1	78.4	76.8	78.7	77.6
98.4	Ar2	87.9	80.6	81.1	84.4	84.9	85.4	83.3	76.0	76.6	74.9	75.7	75.7	74.1	77.4	76.8	75.7	76.0
100	98.4	Au1	85.2	81.1	86.0	87.9	87.1	84.6	77.1	77.6	76.8	76.6	77.6	76.6	77.4	78.2	77.1	78.2
99.1	97.6	99.1	G1	84.4	84.1	81.9	82.5	83.0	78.4	77.9	77.6	78.2	77.9	76.6	77.6	78.2	77.4	76.6
98.4	96.7	98.4	99.2	G2	81.7	83.8	83.0	81.9	78.2	77.6	78.2	77.9	78.4	77.1	77.9	77.6	78.7	76.8
97.6	95.9	97.6	96.7	96.7	Iu1	84.6	86.8	84.9	77.6	77.6	77.1	77.4	77.4	76.3	77.6	77.4	77.6	78.4
99.2	97.6	99.2	98.4	97.6	96.7	US-1	91.9	90.8	75.5	74.9	75.2	75.2	75.7	75.2	76.6	75.5	76.0	76.6
100	98.4	100	99.2	98.4	97.6	99.2	US-2	89.9	75.2	75.4	75.2	75.4	76.0	74.9	75.7	75.7	76.3	77.6
97.6	95.9	97.6	96.7	95.9	95.1	96.7	S1	97.6	76.6	76.6	74.7	74.9	74.4	73.6	75.2	74.9	75.2	75.7
91.1	90.2	91.1	90.2	90.2	92.7	91.9	91.1	88.6	B1	98.7	94.6	94.6	94.9	94.3	94.3	96.0	94.6	79.0
91.1	90.2	91.1	90.2	90.2	92.7	90.2	91.1	89.6	98.4	B2	93.8	93.8	94.1	93.5	93.5	95.1	93.8	78.4
89.4	88.6	89.4	88.6	88.6	91.1	90.2	89.4	87.0	98.4	96.7	C1	97.8	98.1	96.8	92.7	91.6	97.3	79.8
90.2	89.4	90.2	89.4	89.4	91.9	91.1	90.2	87.8	99.2	97.6	99.2	C2	98.7	97.6	93.8	91.6	98.4	79.5
90.2	89.4	90.2	89.4	89.4	91.9	91.1	90.2	87.8	99.2	97.6	99.2	100	C3	97.3	93.5	91.9	98.1	79.5
89.4	88.6	89.4	88.6	88.6	91.9	90.2	89.4	87.0	98.4	96.7	99.2	99.2	99.2	C4	93.0	91.9	97.0	78.7
90.2	89.4	90.2	89.4	89.4	91.9	91.1	90.2	87.8	99.2	97.6	97.6	98.4	98.4	97.6	I1	91.4	93.3	79.2
88.6	87.8	88.6	87.8	87.8	90.2	89.4	88.6	86.2	97.6	95.9	95.9	96.7	96.7	95.9	96.7	I2	91.6	78.4
90.2	89.4	90.2	89.4	89.4	91.9	91.1	90.2	87.8	99.2	97.6	99.2	100	100	99.2	98.4	96.7	P1	78.7
92.7	91.1	92.7	91.9	91.9	94.3	93.5	92.7	90.2	95.9	95.1	94.3	95.1	95.1	94.3	95.1	93.5	95.1	M1
Amino Acid Identity																		

TABLE 57
Nucleotide and deduced amino acid identity between isolates of HEV over 148 base (49 amino acid)* ORF 2 fragment

Nucleotide Identity														
Ar1	91.8	87.8	81.6	82.7	83.7	80.6	82.7	87.8	80.6	80.6	80.6	80.6	80.6	80.6
100	Ar2	88.5	83.8	86.5	87.2	85.8	85.1	90.5	85.8	85.1	83.8	85.1	85.1	85.1
100	100	Aut	83.1	88.5	89.2	83.1	85.8	87.8	85.1	83.8	83.1	83.8	83.8	83.1
100	100	100	G1	87.2	87.8	84.5	85.1	85.1	84.5	82.4	82.4	83.1	83.1	82.4
100	100	100	100	G2	83.1	82.4	85.1	87.8	85.1	84.5	82.4	83.8	83.8	83.1
100	100	100	100	100	It1	87.8	85.8	85.8	83.8	83.1	82.4	83.1	83.1	82.4
96.9	98.0	98.0	98.0	98.0	98.0	US-1	93.9	90.5	79.0	78.4	76.4	77.0	77.0	78.4
96.9	98.0	98.0	98.0	98.0	98.0	100	US-2	91.2	82.4	80.4	79.7	80.4	80.4	81.8
96.9	98.0	98.0	98.0	98.0	98.0	100	100	SI	83.8	84.5	82.4	83.1	83.1	83.8
96.9	98.0	98.0	98.0	98.0	98.0	95.9	95.9	95.9	B1	98	94.6	95.3	95.3	93.9
96.9	95.9	95.9	95.9	95.9	95.9	93.9	93.9	93.9	98.0	B2	93.9	94.6	94.6	93.2
96.9	98.0	98.0	98.0	98.0	98.0	95.9	95.9	95.9	100	98.0	C1	98.0	98.0	96.6
96.9	98.0	98.0	98.0	98.0	98.0	95.9	95.9	95.9	100	98.0	100	C2	100	92.6
96.9	98.0	98.0	98.0	98.0	98.0	95.9	95.9	95.9	100	98.0	100	100	C3	92.6
96.9	98.0	98.0	98.0	98.0	98.0	95.9	95.9	95.9	100	98.0	100	100	C4	98.6
96.9	98.0	98.0	98.0	98.0	98.0	95.9	95.9	95.9	100	98.0	100	100	100	97.3
96.9	98.0	98.0	98.0	98.0	98.0	95.9	95.9	95.9	100	98.0	100	100	100	95.9
93.8	95.9	95.9	95.9	95.9	95.9	93.9	93.9	93.9	98.0	95.9	98.0	98	98.0	I2
96.9	98.0	98.0	98.0	98.0	98.0	95.9	95.9	95.9	100	98.0	100	100	100	P1
96.9	98.0	98.0	98.0	98.0	98.0	95.9	95.9	95.9	95.9	93.9	95.9	95.9	95.9	MI

Amino Acid Identity

* Over 98 base (32 amino acid) fragment for Ar1

Comparisons of the ORF 1 and ORF 2 amplified sequences indicate that the isolates from the two patients from Argentina are quite distinct from each other, exhibiting 88.4 % and 91.8 % nucleotide sequence identity over these regions of ORF 1 and ORF 2, respectively. The value for ORF1 is lower than the lowest percent nucleotide identities between any Burmese-like isolates, which is 91.4%. for ORF 1. However for ORF2, the nucleotide identity of 91.8% between the two isolates from Argentina is in the range observed for identities between the Burmese-like isolates and ORF 2, which may be due to the shorter length of the fragment.

Phylogenetic analyses were performed as described in Example 7. The relative evolutionary distances between the viral sequences analyzed are readily apparent upon inspection of the unrooted phylogenetic trees generated from the pairwise distances, where the branch lengths are proportional to the relative genetic relationships between the isolates. The phylogenetic trees based on alignments of either 371 nucleotides from ORF 1 (Fig. 14), 148 nucleotides from ORF 2 (Fig. 15) which excludes Ar1, or 98 nucleotides from ORF 2 (Fig. 16), which includes Ar1, are quite similar in overall topology. The Burmese-like isolates and the Mexican isolate represent major branches at one end of the tree. The human US isolates form a distinct group distal to the Mexican and Burmese isolates. The swine HEV-like sequence is closely related to the US human isolates. The four European isolates and two Argentine isolates also form branches distal to the Mexican and Burmese isolates. The major branch between the US-type isolates, represented by the US, Greek, Italian, Austrian and Argentine isolates, and the Burmese-like and Mexican isolates is supported by a bootstrap value of 75.7% and greater in all trees.

Example 15: New Degenerate Primers

Degenerate primers derived from consensus oligonucleotide primers for HEV ORF 1 and ORF 2 were designed based on conserved regions between the full length sequences of isolates from Asia, Mexico, US as described in Example 11, as well as isolates from Greece and Italy. The ORF 1 primer is positioned within the methyltransferase region at nucleotides and 473-451 of the Burmese isolate (GenBank accession number M73218), and amplifies a product

417 nucleotides in length when used in combination with HEVConsORF 1-s1, SEQ ID NO:147; as described in Example 11. The new ORF 1 primer combination includes:

HEVConsORF 1-s1; CTGGCATYACTACTGCGYATTGAGC (SEQ ID NO:147); and
HEVConsORF 1N-a1; CCRTCRARRCARTAGGTGCGGTC (SEQ ID NO:255).

The new ORF 2 primer, at positions 6494-6470 of the Burmese isolate, produces a product 197 nucleotides in length when used in combination with HEVConsORF 2-s1; (SEQ ID NO:150); as described in Example 11. The ORF 2 primers include:

HEVConsORF 2-s1; GACAGAATTRATTTTCGTCGGCTGG (SEQ ID NO:150); and
HEVConsORF 2N-a1; CYTGYTCRTGYTGTTTRTCATAATC (SEQ ID NO:256).

For a second round of amplification, internal primers can be used to produce products 287 and 145 nucleotides in length for ORF 1 and ORF 2, respectively, as described in Example 11. The new combination of ORF 1 primers include:

HEVConsORF 1N-s2; CYGCCYTKGCGAATGCTGTGG (SEQ ID NO:257); and
HEVConsORF 1-a2; GGCAGWRTACCARCGCTGAACATC (SEQ ID NO:178).

The ORF 2 primers include:

HEVConsORF 2-s2; GTYGTCTCRGCCAATGGCGAGC (SEQ ID NO:152); and
HEVConsORF 2N-a2; GYTCTRTGYTGRTTTRTCATAATCCTG (SEQ ID NO:258).

PCR reactions contained 2 mM MgCl₂ and 0.5 μM of each oligonucleotide primer as per the manufacturer's instructions (Perkin-Elmer) and amplified using Touch-down PCR as described in Example 5. Amplified products were separated on a 1.5% agarose gel, stained with ethidium bromide, and analyzed for the presence of PCR products of the appropriate size. The primers were used to detect the presence of virus in serum containing HEV as described above and showed a marked increase in sensitivity over previous primers sets used in Example 11. These new primer combinations were found to be more sensitive with a number of different

variants of HEV that included two new isolates from Argentina, Ar1 and Ar2, and a new isolate from Austria, Au1 (see example 14 above), as well as isolates from Greece, G1, and Egypt, Eg46. The results are presented in Table 58 below in which NT represents samples not tested, “-” represents no product band detectable by ethidium bromide staining, “+/-” represents a weak product band detectable by ethidium bromide staining, and “2+”, “3+” and “4+” represent increasing amounts of product as detected by ethidium bromide staining.

TABLE 58

SAMPL E	ORF1				ORF2			
	PCR1		PCR2		PCR1		PCR2	
	Old Set	New Set	Old Set	New Set	Old Set	New Set	Old Set	New Set
Ar 1	-	2+	2+	4+	2+	4+	3+	4+
Ar 2	-	2+	3+	4+	+/-	+/-	-	3+
Au 1	-	2+	3+	4+	-	3+	3+	4+
Eg46	NT	NT	NT	NT	-	3+	3+	4+
G1	-	-	2+	-	3+	3+	3+	4+

Equivalents

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The foregoing embodiments are therefore to be considered in all respects illustrative rather than limiting on the invention described herein. Scope of the invention is thus indicated by the appended claims rather than by the foregoing description, and all changes that come within the meaning and range of equivalency of the claims are intended to be embraced therein.

WHAT IS CLAIMED IS:

1. A method of detecting the presence of a US-type or US-subtype hepatitis E virus (HEV) or a naturally occurring variant thereof in a test sample, the method comprising the steps of:

5 (a) contacting the sample with a binding partner that binds specifically to a marker for said virus, which if present in the sample binds to the binding partner to produce a marker-binding partner complex, and

(b) detecting the presence of said complex, the presence of said complex being indicative of the presence of said virus in the sample.

10 2. The method of claim 1, wherein said marker is an antibody capable of binding said virus.

3. The method of claim 2, wherein said antibody is an immunoglobulin G or an immunoglobulin M.

4. The method of claim 2, wherein said binding partner is an isolated polypeptide chain.

15 5. The method of claim 4, wherein said polypeptide chain is immobilized on a solid support.

6. The method of claim 4, wherein said binding partner is a polypeptide chain selected from the group consisting of SEQ ID NOS:91, 92, and 93, including naturally occurring variants thereof.

20 7. The method of claim 4, wherein said binding partner is a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NO:173 or SEQ ID NO:175.

8. The method of claim 4, where said binding partner is a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NO:174 or SEQ ID NO:176.

9. The method of claim 4, wherein said binding partner is a polypeptide chain selected from the group consisting of SEQ ID NOS:166, 167 and 168, including naturally occurring variants thereof.

10. The method of claim 4, wherein said binding partner is a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:223.

11. The method of claim 4, wherein said binding partner is a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:224.

12. The method of claim 1, wherein said binding partner is an isolated antibody capable of binding specifically to a polypeptide chain selected from the group consisting of SEQ ID NOS:91, 92, 93, 166, 167, and 168, including naturally occurring variants thereof.

13. The method of claim 12, wherein said antibody is a monoclonal antibody.

14. The method of claim 1, wherein said marker is a polypeptide chain.

15. The method of claim 14, wherein said polypeptide chain is selected from the group consisting of SEQ ID NOS:91, 92, and 93, including naturally occurring variants thereof.

16. The method of claim 14, wherein said polypeptide chain comprises the amino acid sequence set forth in SEQ ID NO:173 or SEQ ID NO:175.

17. The method of claim 14, wherein said polypeptide chain comprises the amino acid sequence set forth in SEQ ID NO:174 or SEQ ID NO:176.

18. The method of claim 14, wherein said polypeptide chain is selected from the group consisting of SEQ ID NOS:166, 167, and 168, including naturally occurring variants thereof.

19. The method of claim 14, wherein said polypeptide chain comprises the amino acid sequence set forth in SEQ ID NO:223.

20. The method of claim 14, wherein said polypeptide chain comprises the amino acid sequence set forth in SEQ ID NO:224.

21. The method of claim 1, wherein said marker is a nucleic acid sequence defining at least a portion of a genome of said virus, or a complementary strand thereof.

22. The method of claim 1 wherein said binding partner is an isolated nucleic acid sequence that is capable of hybridizing under specific hybridization conditions to the nucleic acid sequences set forth in SEQ ID NOS:89 and 164.

23. The method of claim 1 wherein said binding partner is selected from the group consisting of SEQ ID NOS:126, 128, 147, 148, 150, 152, 177, 178, 255, 256, 257, and 258.

24. The method of claim 1 wherein said binding partner is an isolated polypeptide chain.

25. The method of claim 1 wherein said test sample is a mammalian cell line.

26. The method of claim 41 wherein said mammalian cell line is a human fetal kidney cell line.

27. A method of detecting the presence of a hepatitis E virus (HEV) in a test sample, the method comprising the steps of:

(a) contacting the sample with a binding partner selected from the group consisting of SEQ ID NOS: 126, 128, 147, 148, 150, 152, 177, 178, 255, 256, 257, and 258 that binds specifically to a marker for said virus, which if present in the sample binds to the binding partner to produce a marker-binding partner complex, and

(b) detecting the presence of said complex, the presence of said complex being indicative of the presence of said virus in the sample.

28. An isolated polypeptide chain comprising the amino acid sequence set forth in SEQ ID NO:173, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176, SEQ ID NO:223 and SEQ ID NO:224.

29. An isolated antibody capable of binding specifically to a polypeptide chain selected from the group consisting of a polypeptide encoded by an ORF 1 sequence of a US-type or a US-subtype HEV, a polypeptide encoded by an ORF 2 sequence of a US-type or a US-subtype HEV, and a polypeptide encoded by an ORF 3 sequence of a US-type or a US-subtype HEV.

30. An isolated antibody capable of binding specifically to a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NO:173, SEQ ID NO:175 or SEQ ID NO:224.

31. An isolated antibody capable of binding specifically to a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NO:174, SEQ ID NO:176 or SEQ ID NO:223.

5 32. The isolated antibody of claim 30, wherein said antibody, under similar conditions, has a lower affinity for a polypeptide chain comprising the amino acid sequence set forth in SEQ ID NO:169 or 171.

10 33. The isolated antibody of claim 31, wherein said antibody, under similar conditions, has a lower affinity for a polypeptide chain comprising the amino acid sequence set forth SEQ ID NO:170 or 172.

34. The isolated antibody of claim 29 further comprising a detectable moiety.

35. An isolated nucleic acid sequence defining at least a portion of an ORF 1, ORF 2 or ORF 3 sequence of a US-type or US-subtype hepatitis E virus, or a sequence complementary thereto.

15 36. An isolated nucleic acid sequence capable of hybridizing under specific hybridization conditions to the nucleotide sequence set forth in SEQ ID NOS:89 and 164.

37. A vector comprising the isolated nucleic acid sequence of claim 35.

38. A host cell containing the vector of claim 37.

20 39. A method of immunizing a mammal against a US-type or US-subtype HEV, the method comprising administering to the mammal the polypeptide of claim 28 in an amount sufficient to stimulate the production of an antibody capable of binding specifically to the US-type or US-subtype hepatitis E virus.

40. A method of immunizing a mammal against a US-type or US-subtype HEV 1, the method comprising administering to said mammal the antibody of claim 29 in an amount
25 sufficient to immunize said mammal against the US-type or US-subtype hepatitis E virus.

41. A method of immunizing a mammal against a US-type or US-subtype HEV 1, the method comprising administering to said mammal the antibody of claim 30 in an amount sufficient to immunize said mammal against the US-type or US-subtype hepatitis E virus.

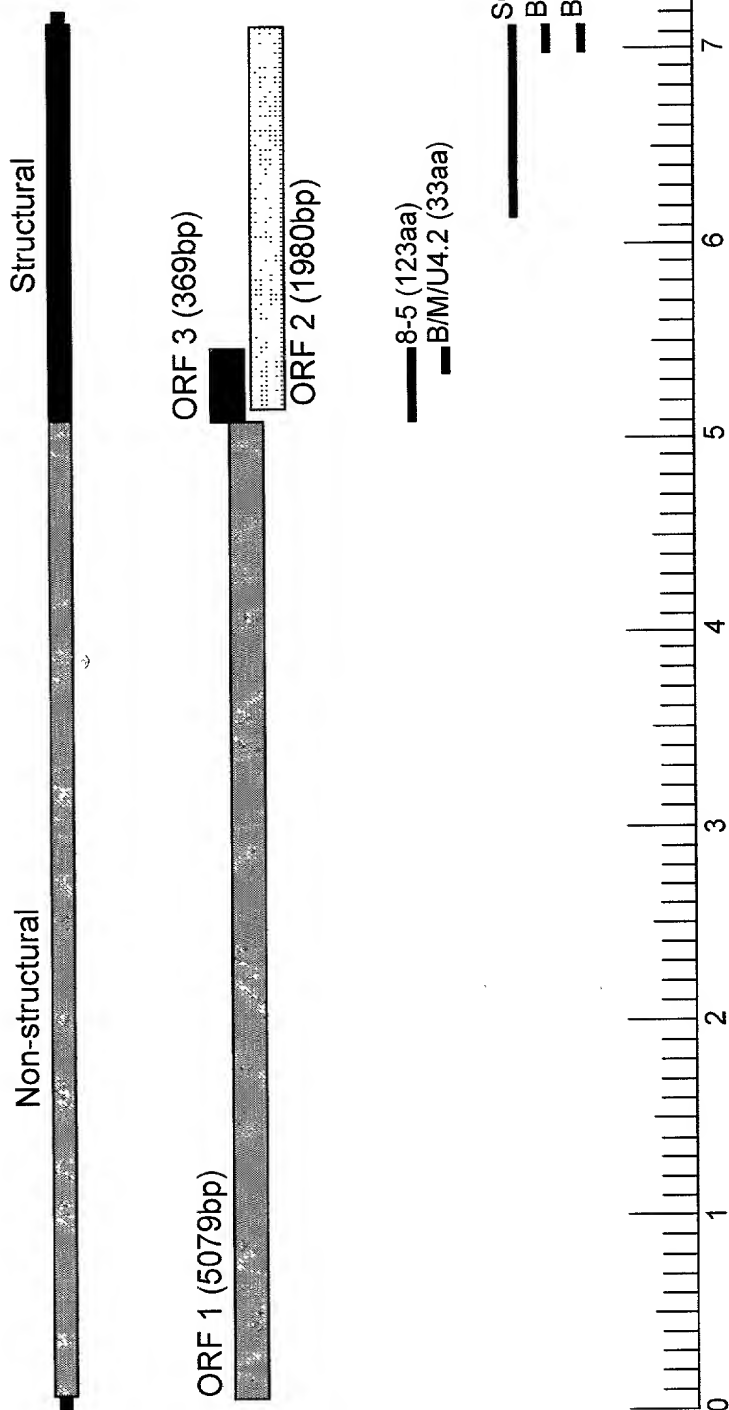
42. A method of immunizing a mammal against a US-type or US-subtype HEV 1, the
5 method comprising administering to said mammal the antibody of claim 31 in an amount sufficient to immunize said mammal against the US-type or US-subtype hepatitis E virus.

43. A method of immunizing a mammal against a US-type or US-subtype HEV, the method comprising administering to said mammal the host cell of claim 38 in an amount sufficient to immunize said mammal against the US-type or US-subtype hepatitis E virus.

ABSTRACT OF THE DISCLOSURE

Disclosed herein are methods and compositions for detecting the presence in a sample of a US-type or a US-subtype hepatitis E virus, including naturally occurring variants thereof. In particular, the invention provides nucleic acid sequences corresponding to the genome of the
5 US-type or US-subtype hepatitis E virus, amino acid sequences, including epitope sequences, encoded by the genomes of such viruses, and antibodies that bind specifically to such amino acid sequences. The invention further provides methods and compositions for immunizing individuals against infection by, or for treating individuals already infected with such a virus.

HEV Genome



1/22

Figure 1

2/22

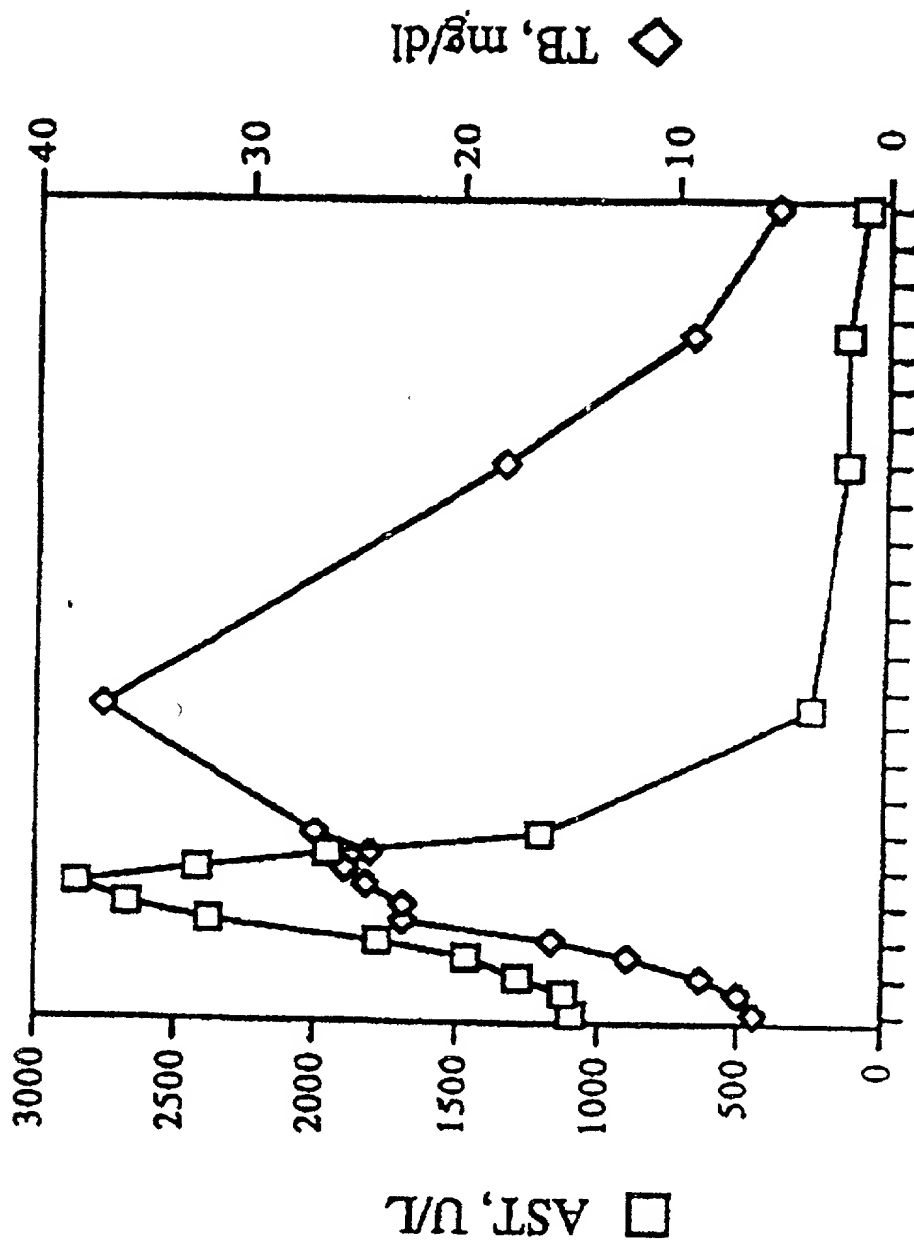


Figure 2

HEV US-1 Genome Extension

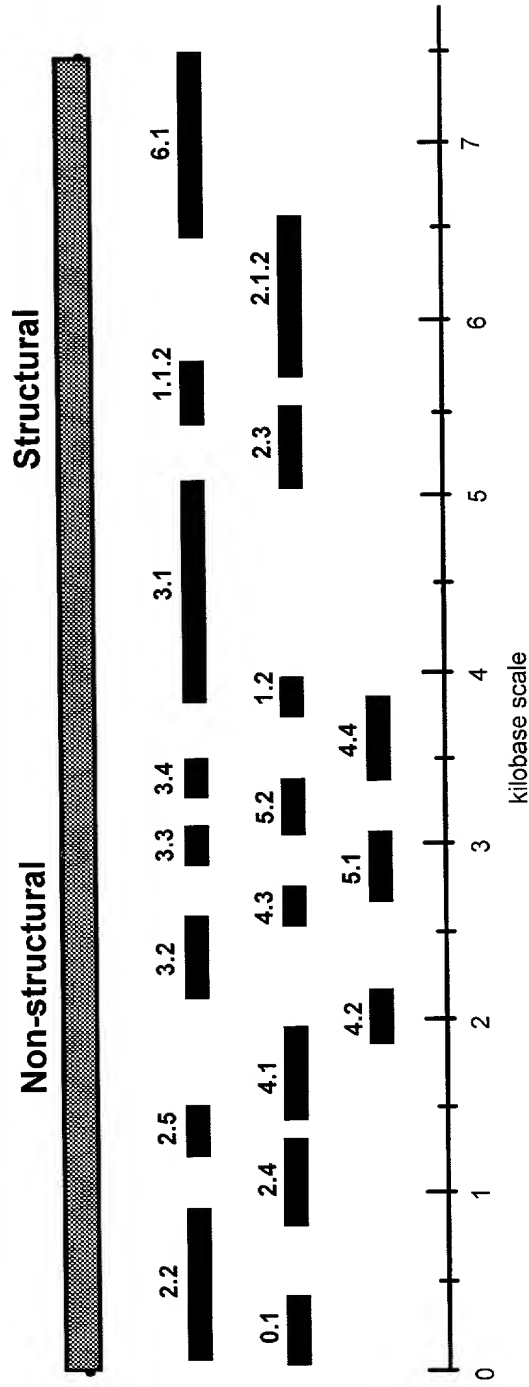


Figure 3

4/22



5/22

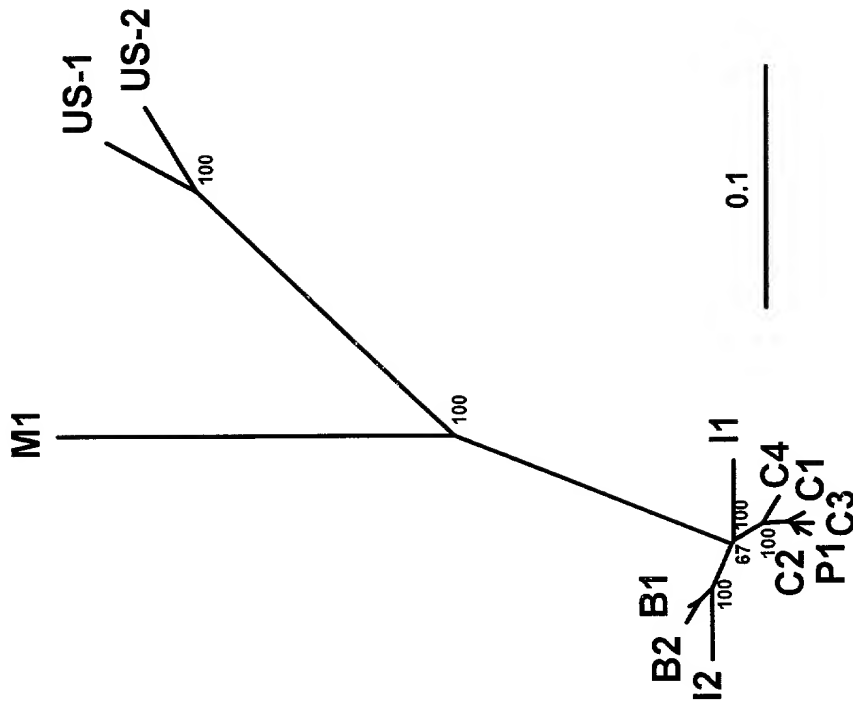


Figure 5

6/22

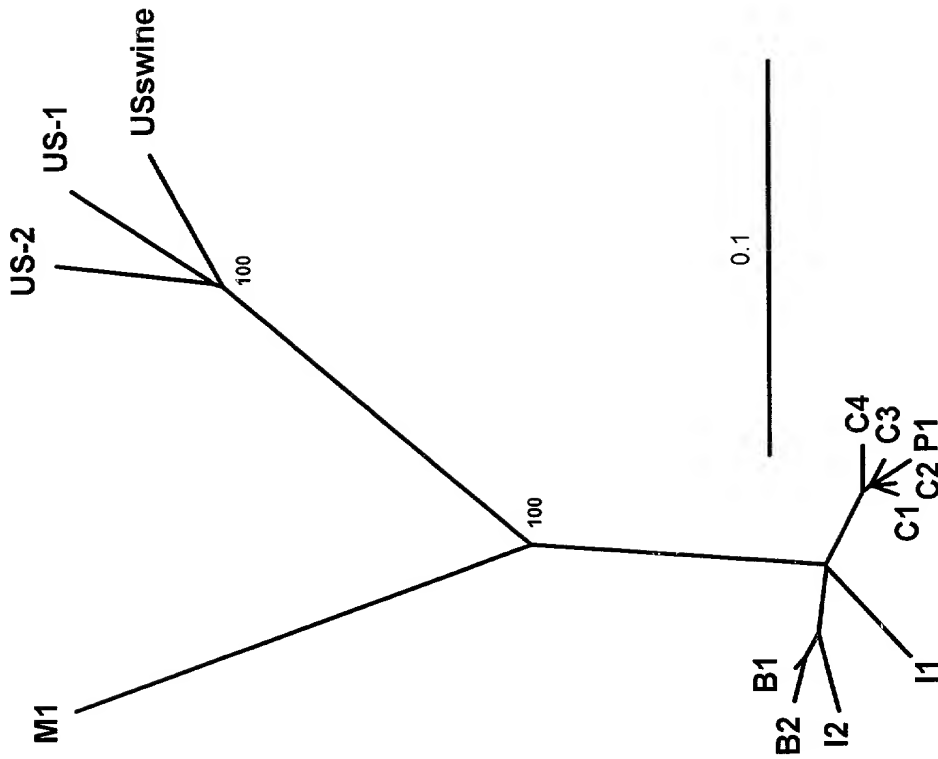


Figure 6

7/22

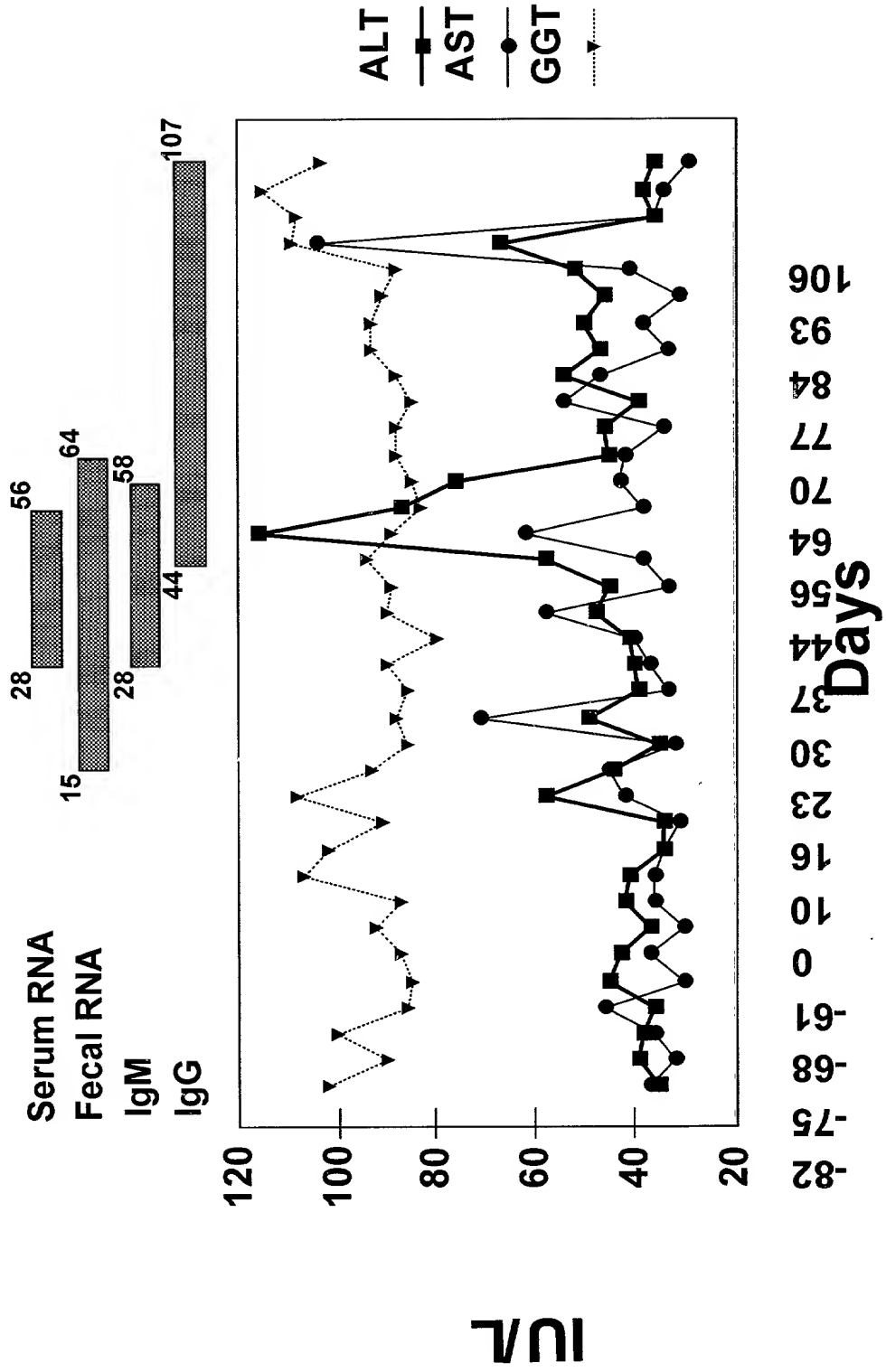


Figure 7

Extension of Z12

8/22

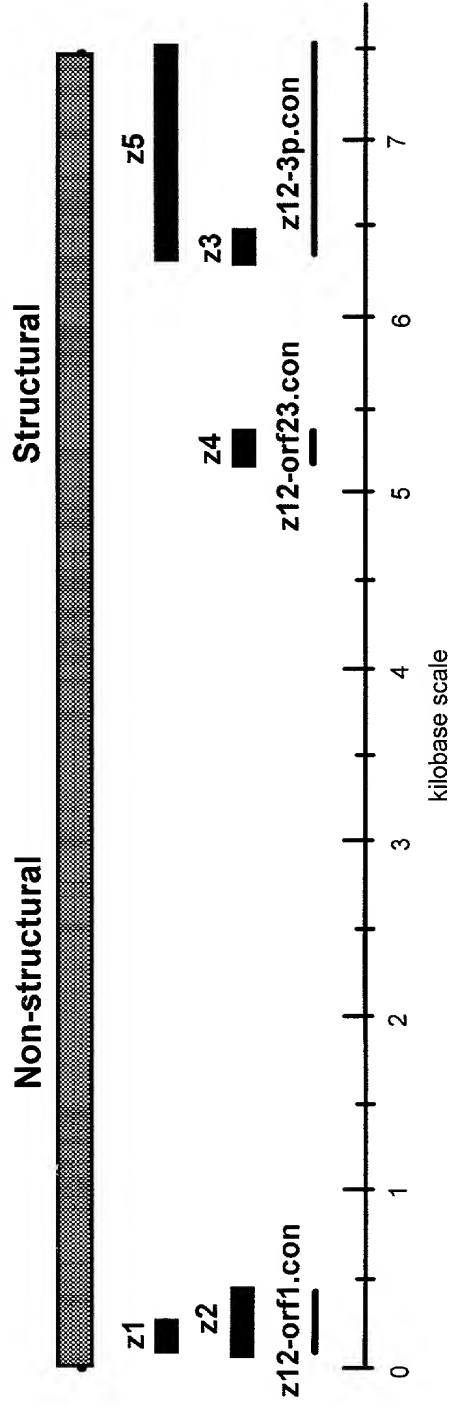


Figure 8

FIGURE 9A

5p.pile{hpesvp}	51	GGCTCCTGGC	ATCACTACTG	CTATTGAGCA	GGCTGCTCTA	GCAGCGGCCA	100
5p.pile{hpeuigh}		GGCTCCTGGC	ATCACTACTG	CTATTGAGCA	GGCTGCTCTA	GCAGCGGCCA	
5p.pile{hpea}		GGCTCCTGGC	ATCACTACTG	CTATTGAGCA	GGCTGCTCTA	GCAGCGGCCA	
5p.pile{840455p}		~~~CCTGGC	ATTACTACTG	CCATTGAGCA	GGCTGCTCTG	GCTGCGGCCA	
5p.pile{hpenssp}		GGCTCCTGGC	ATCACTACTG	CTATTGAGCA	AGCAGCTCTA	GCAGCGGCCA	
Consensus		----CCTGGC	AT-ACTACTG	C-ATTGAGCA	-GC-GCTCT-	GC-GCGGCCA	
5p.pile{hpesvp}	101	ACTCTGCCCT	GCGAATGCT	GTGGTAGTTA	GGCCTTTTCT	CTCTCACCAG	150
5p.pile{hpeuigh}		ATTCTGCCCT	TGCGAATGCT	GTGGTAGTTA	GGCCTTTTCT	CTCTCACCAG	
5p.pile{hpea}		ACTCTGCCCT	TGCGAATGCT	GTGGTAGTTA	GGCCTTTTCT	CTCTCACCAG	
5p.pile{840455p}		ATTCTGCCCT	GCGAATGCT	GTGGTGGTTC	GGCCGTTTTT	ATCTCGCGTG	
5p.pile{hpenssp}		ACTCCGCCCT	TGCGAATGCT	GTGGTGGTCC	GGCCTTTTCT	TTCCCATCAG	
Consensus		A-TC-GCC-T	-GCGAATGCT	GTGGT-GT--	GGCC-TT--T	-TC-C----G	
5p.pile{hpesvp}	151	CAGATTGAGA	TCCTCATTAA	CCTAATGCAA	CCTCGCCAGC	TTGTTTTCCG	200
5p.pile{hpeuigh}		CAGATTGAGA	TCCTTATTAA	CCTAATGCAA	CCTCGCCAGC	TTGTTTTCCG	
5p.pile{hpea}		CAGATTGAGA	TCCTTATTAA	CCTAATGCAA	CCTCGCCAGC	TTGTTTTCCG	
5p.pile{840455p}		CAAACCGAGA	TTCTTATTAA	TTTGATGCAA	CCCCGGCAGT	TGGTTTTCCG	
5p.pile{hpenssp}		CAGGTTGAGA	TCCTTATAAA	TCTCATGCAA	CCTCGGCAGC	TGGTGTTCG	
Consensus		CA----GAGA	T-CT-AT-AA	--T-ATGCAA	CC-CG-CAG-	T-GT-TT-CG	

FIGURE 9B

5p.pile{hpesvp}	201	CCCCGAGGTT	TTCTGGAATC	ATCCCATCCA	GCGTGTTCATC	CATAACGAGC	250
5p.pile{hpeuigh}		CCCCGAGGTT	TTCTGGAACC	ACCCCATCCA	GCGTGTTCATC	CATAATGAGC	
5p.pile{hpea}		CCCCGAGGTT	TTCTGGAACC	ATCCCATCCA	GCGTGTTCATC	CATAATGAGC	
5p.pile{840455p}		CCCTGAGGTA	CTTTGGAATC	ACCCATATCCA	GCGGGTTATA	CATAATGAAT	
5p.pile{hpenssp}		TCCTGAGGTT	TTTTTGAATC	ACCCGATTCA	ACGTGTTATA	CATAATGAGC	
Consensus		-CC-GAGGT-	-T-TGGAA-C	A-CC-AT-CA	-CG-GT-AT-	CATAA-GA--	
5p.pile{hpesvp}	251	TGGAGCTTTA	CTGCCGCGCC	CGCTCCGGCC	GCTGTCTTGA	AATTGGCGCC	300
5p.pile{hpeuigh}		TGGAGCTTTA	CTGTCGCGCC	CGCTCCGGCC	GCTGCCCTTGA	AATTGGTGCC	
5p.pile{hpea}		TGGAGCTTTA	CTGTCGCGCC	CGCTCCGGCC	GCTGCCCTCGA	AATTGGTGCC	
5p.pile{840455p}		TAGAACAGTA	CTGCCGGGCT	CGGGCTGGTC	GTTGCTTGA	GTTGGAGCT	
5p.pile{hpenssp}		TTGAGCAGTA	TTGCCGTGCT	CGCTCGGGTC	GCTGCCCTTGA	GATTGGAGCC	
Consensus		T-GA-C--TA	-TG-CG-GC-	CG--C-GG-C	G-TG--T-GA	--TTGG-GC-	
5p.pile{hpesvp}	301	CATCCCCGCT	CAATAAATGA	TAATCCTAAT	GTGGTCCACC	GCTGCTTCCT	350
5p.pile{hpeuigh}		CACCTCGCT	CAATAAACGA	CAATCCTAAT	GTGGTCCACC	GCTGCTTCCT	
5p.pile{hpea}		CACCCCCGCT	CAATAAATGA	CAATCCTAAT	GTGGTCCACC	GTTGCTTCCT	
5p.pile{840455p}		CACCCAAGAT	CCATTAAATGA	CAACCCCAAC	GTTCTGCATC	GGTGTTCCT	
5p.pile{hpenssp}		CACCCACGCT	CCATTAAATGA	TAATCCTAAT	GTCCCTCCATC	GCTGCTTTCT	
Consensus		CA-CC--G-T	C-AT-AA-GA	-AA-CC-AA-	GT--T-CA-C	G-TG-TT-CT	

FIGURE 9C

5p.pile{hpesvp}	351	CCGCCCTGTT	GGGCGTGATG	TTCAGCGCTG	GTATACTGCT	400	CCCACTCGCG
5p.pile{hpeuigh}		CCGCCCTGCC	GGGCGTGATG	TTCAGCGTTG	GTATACTGCT		CCTACCCGCG
5p.pile{hpea}		CCGTCCCTGCC	GGGCGTGATG	TTCAGCGTTG	GTATACTGCC		CCTACCCGCG
5p.pile{840455p}		TAGACCGGTT	GGCCGAGATG	TTCAGCGCTG	GTACTCTGCC		CCACCCGCG
5p.pile{hpenssp}		CCACCCCGTC	GGCCGGGATG	TTCAGCGCTG	GTACACAGCC		CCGACTAGGG
Consensus		----CC-G--	GG-CG-GATG	TTCAGCG-TG	GTA--C-GC-		CC-AC--G-G
5p.pile{hpesvp}	401	GGCCGGCTGC	TAAATTGCCGG	CGTTCCGCGC	TGCGCGGGCT	450	TCCCGCTGCT
5p.pile{hpeuigh}		GGCCGGCTGC	TAAATTGCCGG	GGTTCCGCAC	TGCGCGGGCT		CCCCGCTGCT
5p.pile{hpea}		GGCCGGCTGC	TAAATTGCCGG	CGTTCCGCGC	TGCGCGGGCT		CCCCGCTGCT
5p.pile{840455p}		GCCCTGcGGc	TAAATTGCCGG	cGcTccGGT	TGCGTGGTCT		CCCCCCCCGCT
5p.pile{hpenssp}		GACCTGCGGC	GAACTGTcGC	CGCTCGGCAC	TTCGTGGTCT		GCCACCAGCC
Consensus		G-CC-GC-GC	-AA-TG-CG-	-G-TC-GC--	T-CG-GG-CT		-CC--C-GC-
5p.pile{hpesvp}	451	GACCGCACTT	ACTGCCTCGA	CGGGTTTCT	GGCTGTAAC	500	TTCCCGCCGA
5p.pile{hpeuigh}		GACCGCACTT	ACTGCCTCGA	CGGGTTTCT	GGCTGTAAC		TTCCCGCCGA
5p.pile{hpea}		GACCGCACTT	ACTGCCTCGA	CGGGTTTCT	GGCTGTAAC		TTCCCGCCGA
5p.pile{840455p}		GACCGCACTT	ACTGCCTTGA	TGGATTCTCC	CGTTGTGCTT		TTGCTGCAGA
5p.pile{hpenssp}		GACCGCACTT	ACTGTTTGA	TGGCTTTGCC	GGCTGCCGTT		TTGCCGCCGA
Consensus		GACCGCACTT	ACTG--T-GA	-GG-TT--C-	-G-TG----T		TT-C-GC-GA

FIGURE 9D

3p.pile{hpea}	1451	ACTGAGTCAG	TGAAGCCAGT	GCTTGACCTG	ACAAATTCAA	TTCTGTGTCG	1500
3p.pile{hpeuigh}		ACTGAGTCGG	TGAAGCCAGT	GCTCGACTTG	ACAAATTCAA	TCCTGTGTCG	
3p.pile{hpesvp}		ACTGAGTCAG	TAAAACCAGT	GCTCGACTTG	ACAAATTCAA	TCTTGTGTCG	
3p.pile{hpenssp}		ACAGAGTCTG	TTAAGCCTAT	ACTTGACCTT	ACACACTCAA	TTATGCACCG	
3p.pile{840453p}		ACAGAGACTA	TTAAACCTGT	ACTTGATCTC	ACAAATTCCA	TCATACAGCG	
Consensus		AC-GAG-C--	T-AA-CC--T	-CT-GA--T-	ACA-A-TC-A	T--T----CG	
3p.pile{hpea}	1501	GGTGGAATGA	ATAACATGTC	TTTTGCTGCG	CCCATGGGTT	CGCGACCATG	1550
3p.pile{hpeuigh}		GGTGGAATGA	ATAACATGTC	TTTTGCTGCG	CCCATGGGTT	GGCGACCATG	
3p.pile{hpesvp}		GGTGGAATGA	ATAACATGTC	TTTTGCTGCG	CCCATGGGTT	CGCGACCATG	
3p.pile{hpenssp}		GTCTGAATGA	ATAACATGTC	GTTTGTGTCG	CCCATGGGTT	CGCCACCATG	
3p.pile{840453p}		GGTGGAATGA	ATAACATGTC	TTTTGCAATCG	CCCATGGGAT	C...ACCATG	
Consensus		G---GAATGA	ATAACATGT-	-TTTGC--CG	CCCATGGG-T	----ACCATG	
3p.pile{hpea}	1551	CGCCCTCGGC	CTATTTTGCT	GTTGCTCCTC	ATGTTTCTGC	CTATGCTGCC	1600
3p.pile{hpeuigh}		CGCCCTCGGC	CTATTTTGCT	GTTGCTCCTC	ATGTTTCTGC	CTATGCTGCC	
3p.pile{hpesvp}		CGCCCTCGGC	CTATTTTGCT	GCTGCTCCTC	ATGTTTCTGC	CTATGCTGCC	
3p.pile{hpenssp}		CGCCCTAGGC	CTCTTTTGCT	GTTGTTCCCTC	TTGTTTCTGC	CTATGTTGCC	
3p.pile{840453p}		CGCCCTAGGC	CTGTTCTGTT	GTTGTTCCCTC	ATGTTTCTGC	CTATGCTGCC	
Consensus		CGCCCT-GG-	CT-TT-TG-T	G-TG-TCCCTC	-TGTTT-TGC	CTAT--TGCC	

FIGURE 9E

3p.pile{hpea}	1601	CGCGCCACCG	CCCGGTCAGC	CGTCTGGCCG	CCGTCGTGGG	CGGCGCAGCG	1650
3p.pile{hpeuigh}		CGCGCCACCG	CCCGGTCAGC	CGTCTGGCCG	CCGTCGTGGG	CGGCGCAGCG	
3p.pile{hpesvp}		CGCGCCACCG	CCCGGTCAGC	CGTCTGGCCG	CCGTCGTGGG	CGGCGCAGCG	
3p.pile{hpenssp}		CGCGCCACCG	ACCGGTCAGC	CGTCTGGCCG	CCGTCGTGGG	CGGCGCAGCG	
3p.pile{840453p}		CGCGCCACCG	GCCGGTCAGC	CGTCTGGCCG	TCGCCGTGGG	CGGCGCAGCG	
Consensus		CGCGCCACCG	-CCGGTCAGC	CGTCTGGCCG	-CG-CGTGGG	CGGCGCAGCG	
3p.pile{hpea}	1651	GCGGTTCCGG	GCGTGGTTTC	TGGGGTGACC	GGGTTGATTC	TCAGCCCCTTC	1700
3p.pile{hpeuigh}		GCGGTTCCGG	GCGTGGTTTC	TGGGGTGACC	GGGTTGATTC	TCAGCCCCTTC	
3p.pile{hpesvp}		GCGGTTCCGG	GCGTGGTTTC	TGGGGTGACC	GGGTTGATTC	TCAGCCCCTTC	
3p.pile{hpenssp}		GCGGTACCGG	GCGTGGTTTC	TGGGGTGACC	GGGTTGATTC	TCAGCCCCTTC	
3p.pile{840453p}		GCGGTGCCGG	GCGTGGTTTC	TGGAGTGACA	GGGTTGATTC	TCAGCCCCTTC	
Consensus		GCGGT-CCGG	GCGTGGTTTC	TGG-GTGAC-	GGGTTGATTC	TCAGCCCCTTC	
3p.pile{hpea}	1701	GCAATCCCCCT	ATATTTCATCC	AACCAACCCC	TTCGCCCCCCG	ATGTCACCGC	1750
3p.pile{hpeuigh}		GCAATCCCCCT	ATATTTCATCC	AACCAACCCC	TTCGCCCCCCG	ATGTCACCGC	
3p.pile{hpesvp}		GCAATCCCCCT	ATATTTCATCC	AACCAACCCC	TTCGCCCCCCG	ATGTCACCGC	
3p.pile{hpenssp}		GCAATCCCCCT	ATATTTCATCC	AACCAACCCC	TTTGCCCCCAG	ACGTTGCCGC	
3p.pile{840453p}		GCCCTCCCCCT	ATATTTCATCC	AACCAACCCC	TTCGCCGCCG	ATGTCGTTTC	
Consensus		GC--TCCCCCT	ATATTTCATCC	AACCAACCCC	TT-GCC-C-G	A-GT-----C	

FIGURE 9F

3p.pile{hpea}	2651	AGCGCTTACC	CTGTTTAACC	TTGCTGACAC	CCTGCTTGGC	GGTCTACCGA	2700
3p.pile{hpeuigh}		AGCGCTTACC	CTGTTTAACC	TTGCTGACAC	CCTGCTTGGC	GGTCTACCGA	
3p.pile{hpesvp}		AGCCCTCACC	CTGTTCAACC	TTGCTGACAC	TCTGCTTGGC	GGCCTGCCGA	
3p.pile{hpenssp}		AGCTCTAACA	TACTTAACC	TTGCTGACAC	GCTCCTCGGC	GGGCTCCC GA	
3p.pile{840453p}		TGCCCTGACT	CTGTTTAATC	TTGCTGATaC	GCTTCTTGGT	GGTTTACCGA	
Consensus		-GC-CT-AC-	-T--T-AA-C	TTGCTGA-AC	-CT-CT-GG-	GG--T-CCGA	
3p.pile{hpea}	2701	CAGAAATTGAT	TTCGTCGGCT	GGTGGCCAGC	TGTTCTACTC	TCGCCCCGTC	2750
3p.pile{hpeuigh}		CAGAAATTGAT	TTCGTCGGCT	GGTGGCCAGC	TGTTCTACTC	TCGCCCCGTC	
3p.pile{hpesvp}		CAGAAATTGAT	TTCGTCGGCT	GGTGGCCAGC	TGTTCTACTC	CCGCCCCGTT	
3p.pile{hpenssp}		CAGAAATTAAT	TTCGTCGGCT	GGCGGGCAAC	TGTTTTATTTC	CCGCCCCGTT	
3p.pile{840453p}		CAGAAATTGAT	TTCGTCGGCT	GGGGGTCAAC	TGTTTTACTC	CCGCCCCGTT	
Consensus		CAGAAATT-AT	TTCGTCGGCT	GG-GG-CA-C	TGTT-TA-TC	-CG-CC-GT-	
3p.pile{hpea}	2751	GTCTCAGCCA	ATGGCGAGCC	GACTGTTAAG	CTGTATACAT	CTGTGGAGAA	2800
3p.pile{hpeuigh}		GTCTCAGCCA	ATGGCGAGCC	GACTGTTAAG	CTGTATACAT	CTGTAGAGAA	
3p.pile{hpesvp}		GTCTCAGCCA	ATGGCGAGCC	GACTGTTAAG	TTGTATACAT	CTGTAGAGAA	
3p.pile{hpenssp}		GTCTCAGCCA	ATGGCGAGCC	AACCGTGAAG	CTCTATACAT	CAGTGGAGAA	
3p.pile{840453p}		GTCTCgGCCA	ATGGCGAgCC	AACAGTAaAG	TTATACACAT	CTGTTtGAgAA	
Consensus		GTC TC -GCCA	ATGGCGAGCC	-AC-GT-AAG	-T-TA-ACAT	C-GT-GAGAA	

FIGURE 9G

3p.pile{hpea}	2801	TGCTCAGCAG	GATAAGGGTA	TTGCAATCCC	GCATGACATC	GACCTCGGGG	2850
3p.pile{hpeuigh}		TGCTCAGCAG	GATAAGGGTA	TTGCAATCCC	GCATGACATC	GACCTCGGGG	
3p.pile{hpesvp}		TGCTCAGCAG	GATAAGGGTA	TTGCAATCCC	GCATGACATT	GACCTCGGAG	
3p.pile{hpenssp}		TGCTCAGCAG	GATAAGGGTG	TTGCTATCCC	CCACGATATC	GATCTTGGTG	
3p.pile{840453p}		TGCgCAgCAA	gACAAGGGca	TcacCaTTCC	ACACGACATA	gATTTAGGTG	
Consensus		TGC-CAGCA-	GA-AAGGG--	T--C-AT-CC	-CA-GA-AT-	GA--T-GG-G	
3p.pile{hpea}	2851	AATCCCCGTGT	AGTTATTTCAG	GATTATGACA	ACCAACATGA	GCAGGACCGA	2900
3p.pile{hpeuigh}		AATCTCGAGT	TGTTATTTCAG	GATTATGACA	ACCAACATGA	GCAGGACCGG	
3p.pile{hpesvp}		AATCTCGTGT	GGTTATTTCAG	GATTATGATA	ACCAACATGA	ACAAAGATCGG	
3p.pile{hpenssp}		ATTGCGGTGT	GGTCATTTCAG	GATTATGACA	ACCAGCATGA	GCAGGATCGG	
3p.pile{840453p}		ACTCCCCGTGT	GGTTAtCCAG	gattATgATa	Accagcacga	acaAgaTcgA	
Consensus		A-TC-CG-GT	-GT-AT-CAG	GATTATGA-A	ACCA-CA-GA	-CA-GA-CG-	

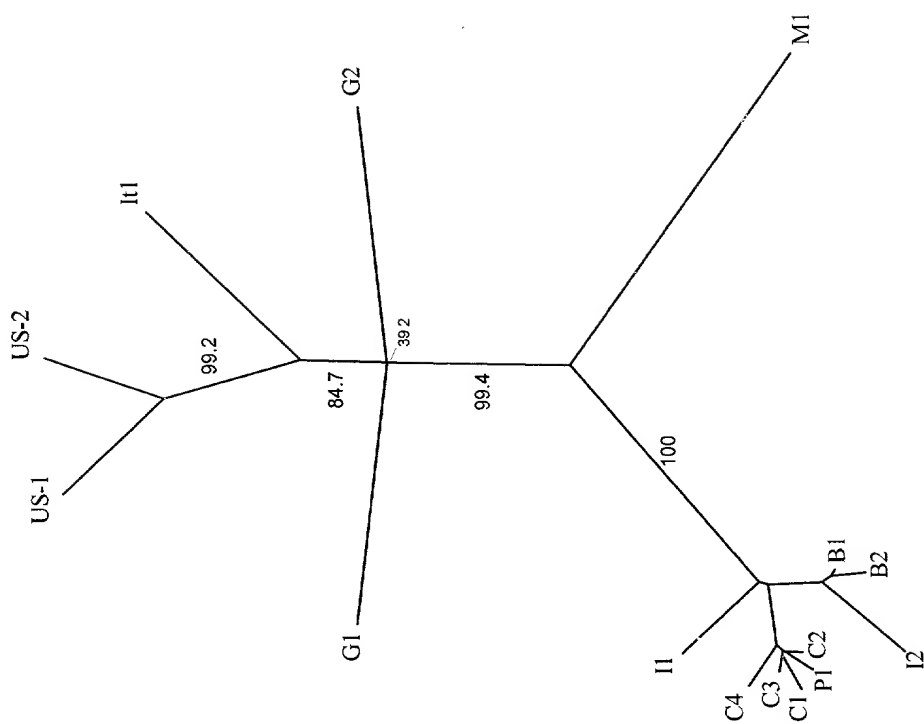


Figure 10

17/22

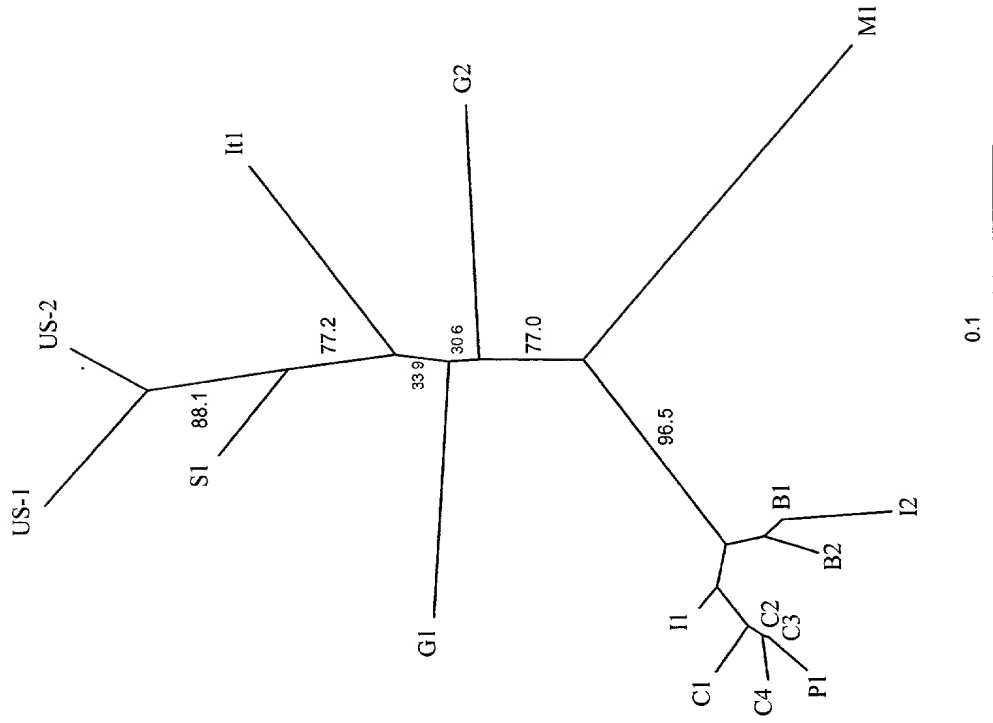
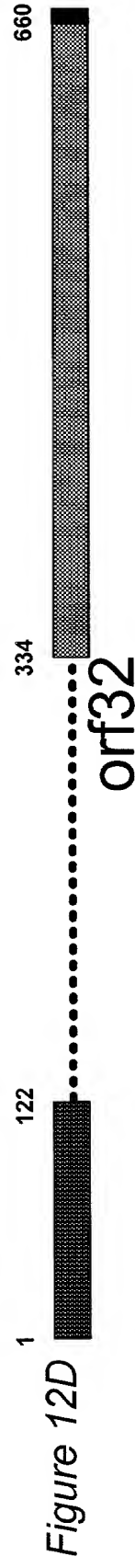
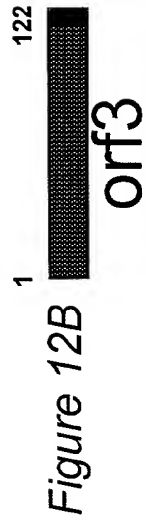
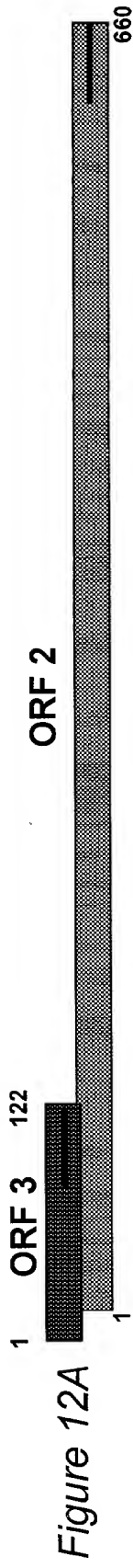


Figure 11



MACAQUE 13906
HEV US-2 ORF 3 CKS - 29

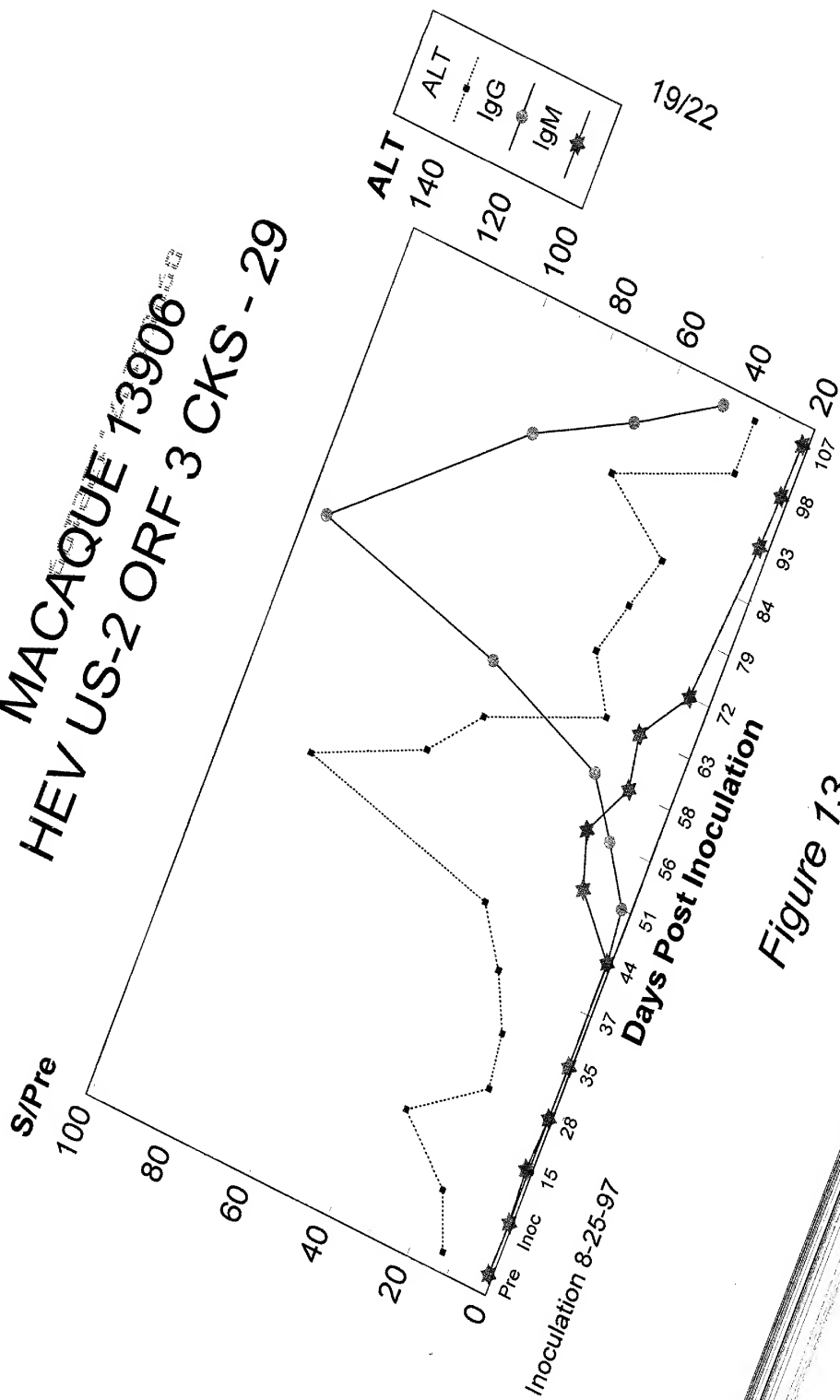


Figure 13

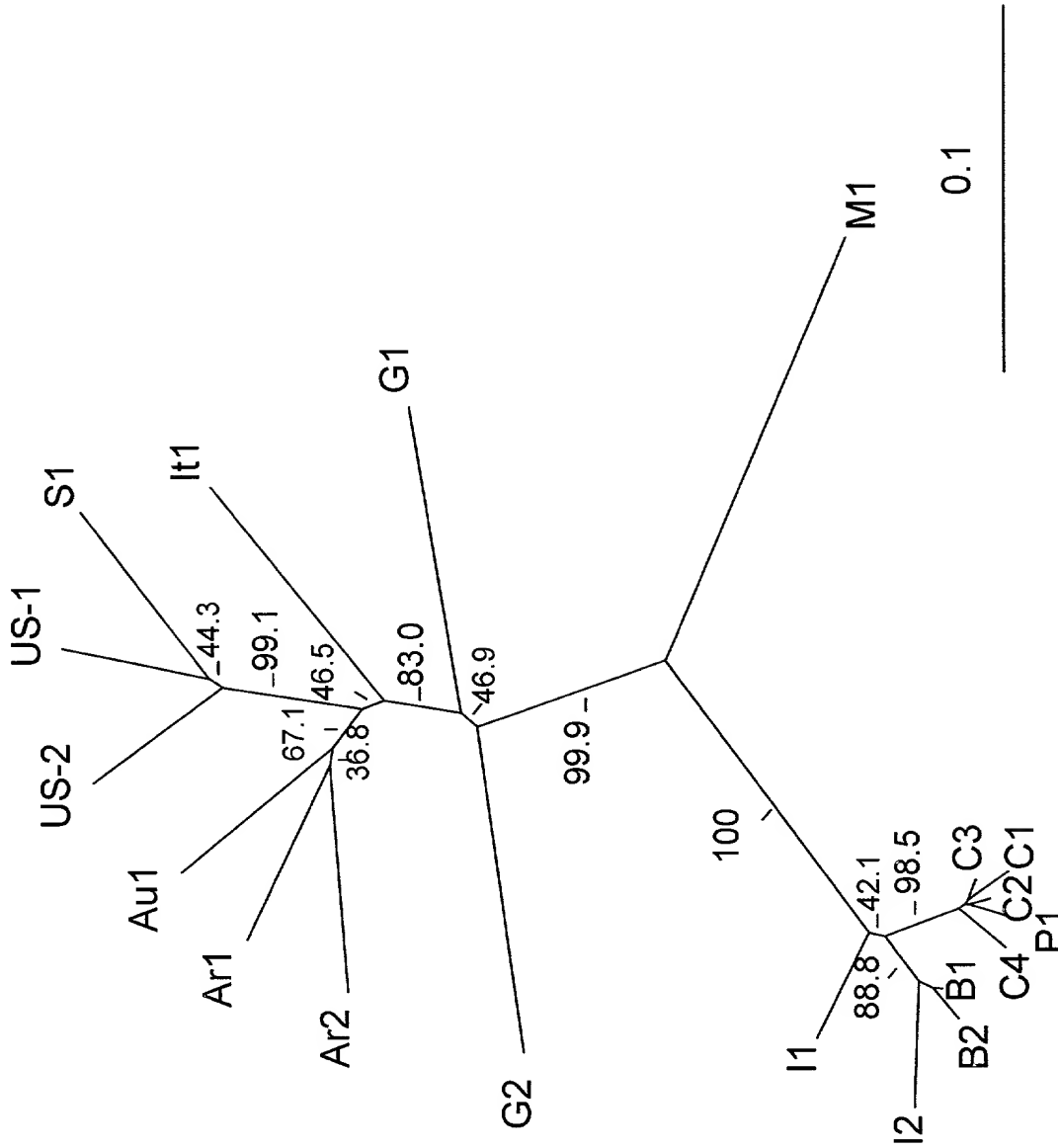


Figure 14

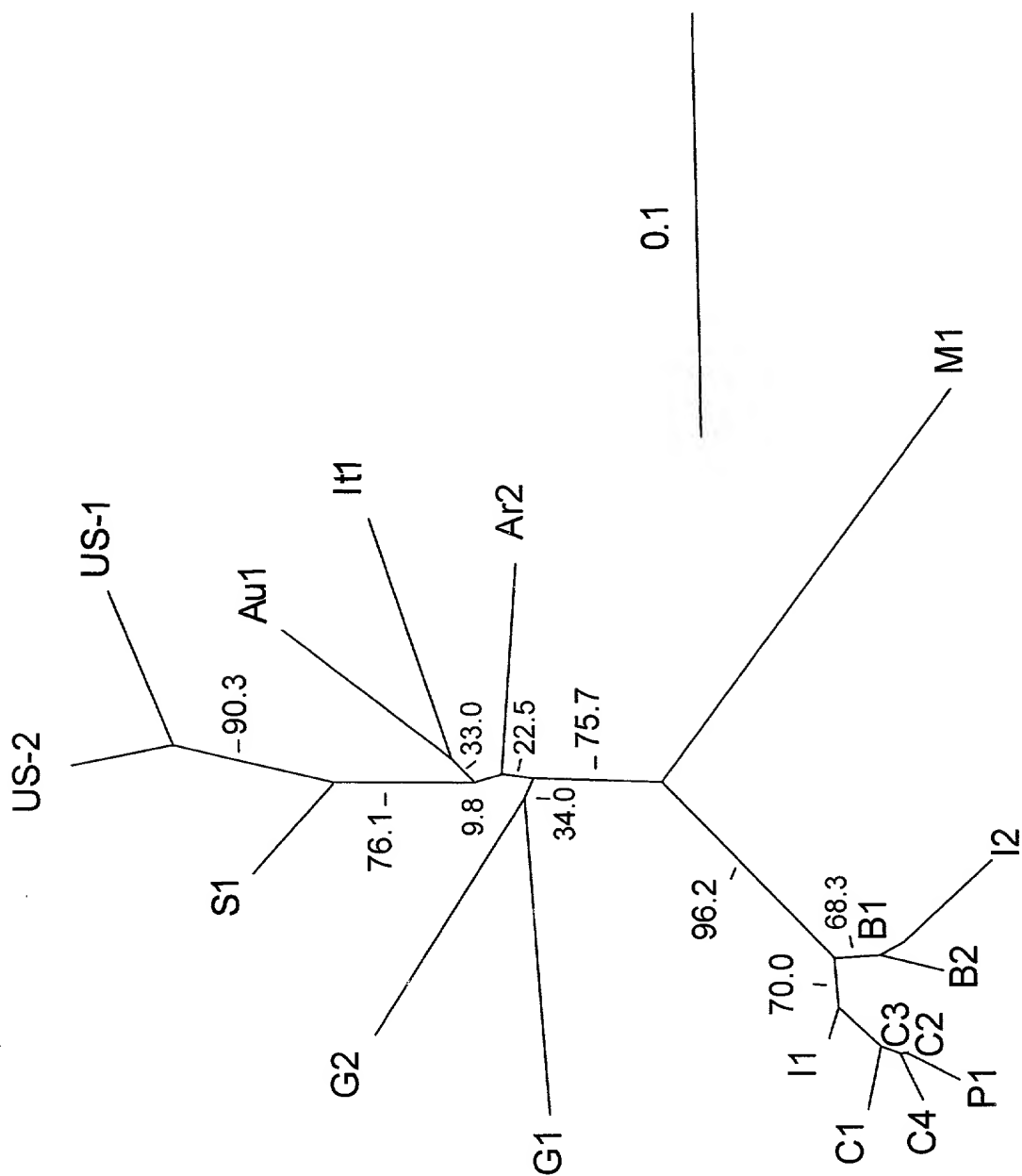


Figure 15

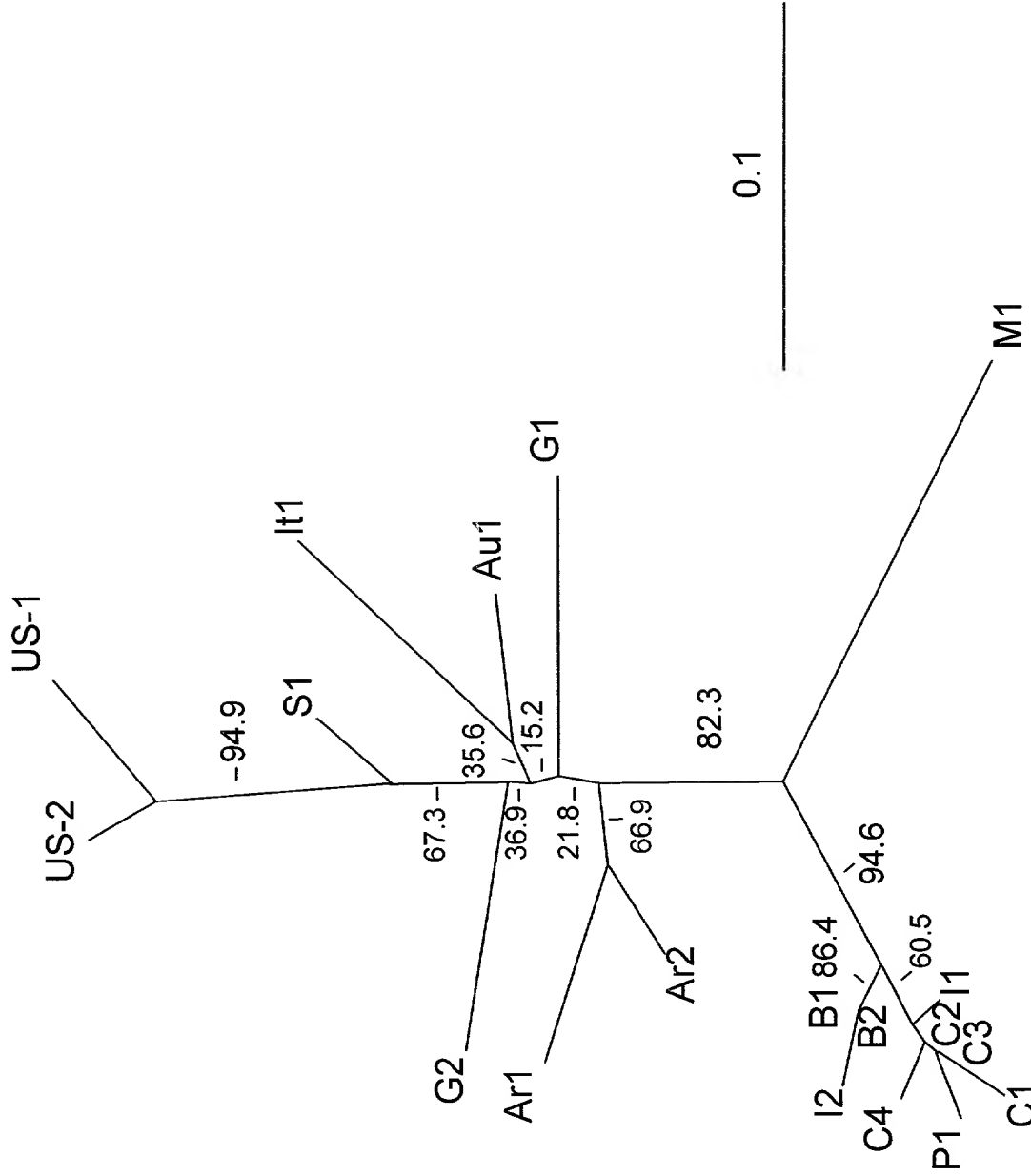


Figure 16

SEQUENCE LISTING

<110> Schlauder, George G
 Erker, James C
 Desai, Suresh M
 Dawson, George J
 Mushawar, Isa K

<120> METHODS AND COMPOSITIONS FOR DETECTING HEPATITIS E
 VIRUS

<130> 6232.US.01

<140>

<141>

<150> US 09/173,141

<151> 1998-10-15

<160> 258

<170> PatentIn Ver. 2.0

<210> 1

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer C375M

<400> 1

ctgaacatcc cggccgac

18

<210> 2

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer A1-350M

<400> 2

agaaagcagc gatggagga

19

<210> 3

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer S1-34M

<400> 3

gccaccagt tcattaaggc t

21

<210> 4
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer A2-320M

<400> 4

tcattaatgg agcgtgggtg

20

<210> 5
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer S2-55M

<400> 5

cctggcatca ctactgctat

20

<210> 6
<211> 18
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer C375

<400> 6

ctgaacatca cgcccaac

18

<210> 7
<211> 19
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer A1-350

<400> 7

aggaagcagc ggtggacca

19

<210> 8
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer S1-34

<400> 8
gccccatcagt ttattaaggc 20

<210> 9
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer A2-320

<400> 9
tcattttattg agcggggatg 20

<210> 10
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer S2-55

<400> 10
cctggcatca ctactgctat 20

<210> 11
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer M1PR6

<400> 11
ccatgttcca caccgtattc cagag 25

<210> 12
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer S4294M

<400> 12
gtgttctacg gggatgctta tgacg 25

<210> 13
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer M1PF6

<400> 13

gactcagtat tctctgctgc cgtgg

25

<210> 14

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer A4556

<400> 14

ggctcaccag aatgcttctt ccaga

25

<210> 15

<211> 342

<212> DNA

<213> Hepatitis E virus

<220>

<223> Clone: USP-15

<400> 15

gcccacagt ttattaaggc tcttgccatt actactgcc a ttgagcaggc tgctctggct 60

gcggcccaatt ctgccttggc gaatgctgtg gtggttcggc cgtttttata tcgcgtgcaa 120

accgagattc ttattaattt gatgcaaccc cggcagttgg ttttcgccc tgaggtactt 180

tggaatcacc ctatccagcg ggttatacat aatgaattag aacagtactg cggggctcgg 240

gctggtcggt gcttgagggt tggagctcac ccaagatcca ttaatgacaa ccccaacggt 300

ctgcatcggt gtttccttag accggtcggg cgtgatgttc ag 342

<210> 16

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer
PA2-5560

<400> 16

taggttatac tgccggcgca

20

<210> 17

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer S1-5287

<400> 17

ttctcagccc ttcgcaatcc

20

<210> 18

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer S2-5310

<400> 18

atattcatcc aaccaacccc

20

<210> 19

<211> 251

<212> DNA

<213> Hepatitis E virus

<220>

<223> Clone b421

<400> 19

atattcatcc aaccaacccc ttgcgcgcgc atgtcgtttc acaaccggg gctggaactc 60

gccctcgaca gccgccccgc cccctcggtt ccgcttggcg tgaccagtcc cagcgcccct 120

ccgttgcccc ccgtcgtcga tctaccccag ctggggctgc gccgctaact gccatatcac 180

cagcccctga tacagtcct gtacctgatg ttgactcacg tgggtctatt ttgcgcgcgc 240

agtataacct a

251

<210> 20

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer
US4.2-69S/20

<400> 20

ttccgcttgg cgtgaccagt

20

<210> 21

<211> 21

<212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer
 US4.4/144s

<400> 21
 gctaactgcc atatcaccag c 21

<210> 22
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer M6417a

<400> 22
 cccttatacct gctgagcatt 20

<210> 23
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer M6371a

<400> 23
 ttggctcgcc attggctgag acaa 24

<210> 24
 <211> 899
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> Clone df-orf2/3

<400> 24
 gctaactgcc atatcaccag ccctgatac agctcctgta cctgatgttg actcacgtgg 60
 tgctattttg cgccggcagt acaatttgtc tacgtccccg cttacatcat ctgttgcttc 120
 tggtaactaat ctggttctct atgctgcccc gctgaaccct ctcttgccctc ttcaggatgg 180
 caccaacact catattatgg ctactgaggc atctaattac gcccagtatc ggggttggtcg 240
 ggctacgatt cgttatcgcc cgttggtgcc aaatgctgtt ggtgggttatg ctatctctat 300
 ttctttctgg cctcaaaact caactacccc tactttctgtt gacatgaatt ctatcacttc 360
 tactgatgtc aggatcttgg tccagccccg tatagcctcc gagttagtca tccctagtga 420

acgccttcac taccgcaacc aaggetggcg ctctgttgag accacgggtg tggccgaaga 480
 ggaggctacc tccggtctgg taatgctttg tattcatggc tccctgtta actcctacac 540
 taatacacct tacaccggtg cattgggggt tcttgatttt gcattagaac ttgaatttag 600
 aaatttgaca cccggaaca ctaacaccog tgtttcccg tatactagca cagcccgcca 660
 ccggctgcg cgcggtgctg atgggaccgc tgagctcacc accacagcag ccacacgctt 720
 catgaaggat ttgcatttta ctggtacgaa cggcggttgg gaggtgggtc gtggtattgc 780
 cctgactctg ttaaatcttg ctgatacgt tcttggtggt ttaccgacag aattgatttc 840
 gtcggctggg ggtcaactgt ttaactcccg cctgttggtc tcagccaatg gcgagccaa 899

<210> 25

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer USP
 3s/20

<400> 25

tggcattact actgccattg 20

<210> 26

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer M902A

<400> 26

atcgatcgga catagacctc 20

<210> 27

<211> 846

<212> DNA

<213> Hepatitis E virus

<220>

<223> Clone df-orf1

<400> 27

tggcattact actgccattg agcaggctgc tctggctgcg gccaatctcg ccttggcgaa 60

tgctgtggtg gtteggccgt ttttatctcg cgtgcaaacc gagattctta ttaatttgat 120

gcaaccccg cagttggttt tccgcctga ggtactttgg aatcaccta tccagcgggt 180

tatacataat gaattagaac agtactgccg ggctcgggct ggtcgttgct tggagggttg 240
 agtcaccca agatccatta atgacaaccc caacgttctg catcgggtgtt tccttagacc 300
 ggttggccga gatgttcagc gctgggtactc tgccccacc cgcggccctg cggctaattg 360
 ccgcgcgtcc gcgttgctg gtctcccccc cgtgaccgc acttactgct ttgatggatt 420
 ctcccgttgt gcttttctg cagagaccgg tgtggctctt tactctctgc atgacctttg 480
 gccagctgat gttgcagagg ctatggcccc ccacgggatr acacgcttgt atgccgact 540
 gcaccttccc cctgagggtg tgetaccacc cggcacctac cacacaacct cgtatctcct 600
 gattcacgac ggcgaccgcg ctgttgtaac ttacgagggc gatactagtgc cgggctataa 660
 tcatgatgtc tccatacttc gtgcgtggat ccgtactaca aaaatagttg gtgatcatcc 720
 gttggtcata gagcgtgtgc gggccattgg atgtcatttt gtgttgctgc tcaccgcagc 780
 ccctgagccg tcacccatgc cttatgttcc ttaccctcgt tcaacggagg tctatgtccg 840
 atcgat 846

<210> 28
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer 3750s

<400> 28
 cttccatcag ttggctgagg agc 23

<210> 29
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer 3900a

<400> 29
 gccatgcggc agtgcaaat gtc 23

<210> 30
 <211> 168
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> Clone HEV 167

<400> 30
 cttccatcag ttggctgagg agctggggcca tgcgccggcc cctgtcgccg ccgtcttgcc 60
 cccttgccct gagcttgagc agggcctgct ctacatgccca caggagctca ctgtgtccga 120
 tagtgtgttg gtttttgagc ttacggacat tgtgcactgc cgcattggc 168

<210> 31
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer 5000s

<400> 31
 ctcgttcata acctgattgg catgc 25

<210> 32
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer
 uf-orf2/3 a3

<400> 32
 ggactggtca cgccaagcgg aac 23

<210> 33
 <211> 424
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> Clone HEV 426

<400> 33
 ctcgttcata acctgattgg catgctgcag accatcgccg atggcaaggc ccactttaca 60
 gagactatta aacctgtact tgatctcaca aattccatca tacagcgggt ggaatgaata 120
 acatgtcttt tgcacgcgcc atgggatcac catgcgccct agggctgttc tgttgttgtt 180
 cctcatgttt ctgcctatgc tgcccgccg accggccggt cagccgtctg gcgctcgccg 240
 tgggcggcgc agcggcggtg ccggcggtgg tttctggagt gacagggttg attctcagcc 300
 cttagccctc ccctatatcc atccaaccaa ccccttcgcc gccgatgtcg tttcacaacc 360
 cggggctgga actcgccctc gacagccgcc ccgccccctc gggtccgctt ggogtgacca 420

gtcc

424

<210> 34

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer 167-s1

<400> 34

tctacatgcc acaggagctc actg

24

<210> 35

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer 426-a3

<400> 35

gatggaattt gtgagatcaa gtacagg

27

<210> 36

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer 167-s2

<400> 36

ctcactgtgt ccgatagtgt gttgg

25

<210> 37

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer 426-a4

<400> 37

ccttgccatc ggcatgggc tgc

23

<210> 38

<211> 1186

<212> DNA

<213> Hepatitis E virus

<220>

<223> Clone HEV 1186

<400> 38

```
ctcactgtgt ccgatagtgt gttggttttt gagcttacgg atatagttca ttgccgcatg 60
gccgctccaa gccagcgaaa ggctgtttct tcaacacttg tggggaggta tggccgtagg 120
acgaaactat atgaggcggc gcattcagat gttcgtgagt ccctagctag gttcatccct 180
actatcgggc ctgttcaggc taccacatgt gagttgtatg agttgggtga ggctatgggtg 240
gagaaaggtc aggacggctc tgcagtctta gagcttgatc tttgtaatcg tgatgtctcg 300
cgcatcacat ttttccaaaa agwctgcaac aagtttacia ctggtgagac catcgcccac 360
ggcaagggtt gccagggtat atcggcctgg agtaagacct tctgcgctct gttcggcccc 420
tggttccgcy ccattgaaaa agaaatattg gccctgctcc cgcctaatat cttttatggc 480
gacgcttatg aggagtcaat ttttgccgcc gctgtgtccg gggcggggtc atgtatggta 540
tttgaaaatg acttttcaga gtttgacagt acccagaata atttctctct tggccttgag 600
tgtgtgggtta tggaggagtg cggcatgcct caatggctaa ttaggttgta ccatctggtt 660
cggctctgct ggattctgca ggcgccgaag gagtctctta agggtttctg gaagaagcat 720
tctggtgagc ctggtaccct tctttggaat accgtctgga atatggcgat tatagcacat 780
tgctatgagt tccgtgactt tctgtgtgct gcctttaagg gtgatgattc ggtggctctc 840
tgtagtgact accgacagag ccgcaatgca gctgccttaa ttgctggctg tgggctcaaa 900
ttgaagggtt attaccgcc taccgggctg tatgctgggg tgggtgggtgg ccccggtttg 960
gggacactgc ccgatgtggt gcgttttctt ggtcgggtgt ctgaaaagaa ttggggcccc 1020
ggcccggaac gtgctgagca gctgcgtctt gctgtctgcy acttccttcg aggggtgacg 1080
aatgttgcyg aggtctgtgt tgatgttgtg tcccggtgtc atggagtcag ccccgggctc 1140
gtacataacc ttattggcat gctgcagacc atcgccgatg gcaagg 1186
```

<210> 39

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer orf1-s2

<400> 39

tcacccatgc cttatgttcc ttacc

25

<210> 40

<211> 22
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer 1300a

<400> 40

ggcggcctgg gatgtaatca cg

22

<210> 41

<211> 460

<212> DNA

<213> Hepatitis E virus

<220>

<223> Clone HEV 459

<400> 41

tcacccatgc cttatgttcc ttaccctcgt tcaacggagg tgtatgtccg gtccatattt 60

ggccctggcg gtcceccatc cttgtttccg tcagcctgct ctactaaatc tactttccat 120

gctgtcccg tgcatatctg ggatcggctc atgctctttg gtgccaccct ggacgatcag 180

gcgttttgct gttcacggct catgacttac ctccgtggta ttagttacaa ggtcactgtc 240

ggcgcgcttg tcgctaata ggggtggaac gcctctgaag acgctcttac tgcartgac 300

actgcagctt atttgactat ttgccatcag cgttatctcc gcaccaggc gatatccaag 360

ggcatgcgcc ggttgggggt tgagcacgcc cagaaattta tcacaagact ctacagttgg 420

ctatttgaga agtctggccg tgattacatc ccaggccgcc 460

<210> 42

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer 459-s2

<400> 42

cagaaattta tcacaagact ctacag

26

<210> 43

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer 1450a

<400> 43
aacactcctg accgagccac ttc

23

<210> 44
<211> 235
<212> DNA
<213> Hepatitis E virus

<220>
<223> Clone HEV 216

<400> 44
cagaaattta tcacaagact ctacagttgg ctatttgaga agtctggccg tgattatata 60
cccggccgcc agcttcagtt ctatgcacag tgccgacggg ggctatctgc aggcttccac 120
ctagacccca gggacttgt ttttgatgag tcagtaccat gccgctgtag gacgtttttg 180
aagaaagttg cgggttaaatt ctgctgtttt atgaagtggc tcggtcagga gtgtt 235

<210> 45
<211> 26
<212> DNA
<213> Hepatitis E virus

<220>
<223> us1 gap-s1

<400> 45
tatagatata acaggttcac ccagcg

26

<210> 46
<211> 24
<212> DNA
<213> Hepatitis E virus

<220>
<223> us1 gap-a0.5

<400> 46
gctgcaagac cctcacgcat gatg

24

<210> 47
<211> 23
<212> DNA
<213> Hepatitis E virus

<220>
<223> us1 gap-s2

<400> 47
cggattatgg ttacaccctg agg

23

<210> 48
 <211> 25
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us1 gap-al

<400> 48
 attcagttgg gtaaaacgct tctgg

25

<210> 49
 <211> 545
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us1-gap

<400> 49
 cggattatgg ttacaccctg aggggttget gggatattttc ccccttttct ccctgggca 60
 tatctgggag tctgcgaacc ccttttgagg ggaggggact ttgtataccc gaacttggtc 120
 aacatctggc ttttctagt atttctcccc cctgaagcg gccgctcctg ctatggctgc 180
 tacccegggg ctgccccatt ctaccccacc tgttagcgat atttgggtgc taccaccgcc 240
 ctgagaggag tttcaggttg atgcagcacc tgtgccccct gccctgacc ctgctggatt 300
 gcccggtccc gttgtgctta cccccccccc cctccccct gtgcataagc catcaatacc 360
 cccgccttcc cgtaaccgct gtctcctcta tacctatcct gacggcgcta aggtgtatgc 420
 agggtcactg tttgaatcag actgtgactg gctgggtaat gctcaaacc cggggccatcg 480
 tcccggaggt ggcctctgcc atgcctttta ccaacgtttt ccagaagcgt tttacccaac 540
 tgaat

545

<210> 50
 <211> 24
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us1-2600s

<400> 50
 gtgctcacca taactgagga cacg

24

<210> 51
 <211> 24

<212> DNA
 <213> Hepatitis E virus

<220>
 <223> us1-2600a

<400> 51
 cgctgcatat gtaacagcaa cagg 24

<210> 52
 <211> 344
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us1-344

<400> 52
 gtgctcacca taactgagga cacggcccggt acggccaacc tggcattgga gattgatgcc 60
 gctacagagg tcggccgtgc ttgtgccggt tgcaccatca gccctggcat tgtgcactat 120
 cagtttaccg ccgggggtccc gggctcgggc aagtcaaggt ccatacaaca gggagatgtc 180
 gatgtggtgg ttgtgcccac ccgggagctt cgtaatagtt ggcgcgcgcg gggttttgcg 240
 gccttcacac cccacacagc ggcccgtgtt actatcggcc gccgcgttgt gattgatgag 300
 gctccatctc tcccgccaca cctgttgctg ttacatatgc agcg 344

<210> 53
 <211> 23
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us1 3200s

<400> 53
 gccgatgtgt gcgagctcat acg 23

<210> 54
 <211> 25
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us1 3200a

<400> 54
 atgattgtgg tctctgtgaa ggtgg 25

<210> 55

<211> 194
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us1-194

<400> 55
 gccgatgtgt gcgagctcat acgcggagcc taccctaaaa tccagaccac gagccgtgtg 60
 ctacgggtccc tgttttggaa tgaaccggcc attggccaga agttgggttyt cacgcaggcg 120
 gcaaaggctg ctaaccctgg tgcgattacg gtccacgaag ctcagggtgc caccttcaca 180
 gagaccacaa tcat 194

<210> 56
 <211> 23
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> HEV216-s1

<400> 56
 cagtaccatg ccgctgtagg acg 23

<210> 57
 <211> 26
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-733a1

<400> 57
 ccattagatg aaatctttac ctgcag 26

<210> 58
 <211> 26
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> HEV216-s2

<400> 58
 gtaggacgtt tttgaagaaa gttgcg 26

<210> 59
 <211> 24
 <212> DNA
 <213> Hepatitis E virus

<220>

<223> us2-733a2

<400> 59

ggtgagctca taagtgaggc tgtg

24

<210> 60

<211> 464

<212> DNA

<213> Hepatitis E virus

<220>

<223> us1-733wb

<400> 60

gtaggacgtt tttgaagaaa gttgcgggta aattctgctg ttttatgcgg tggctcgggc 60

aggagtgtac ctgcttcttg gagccggccg agggtttagt cggcgatcat ggccatgaca 120

acgaggccta tgagggttct gaggtcgacc cggctgaacc tgcacatctt gatgtttctg 180

ggacttacgc cgtccacggg caccagcttg aggccctcta tagggcactt aatgtccac 240

aagatattgc cgctcgagct tcccgaacta cggcaactgt tgagctcggt gcaagtccag 300

accgcttaga gtgccgcacc gtgctcggtg ataagacctt cgggacgacg gtggtcgacg 360

gcgcccctct agaggcgaat ggccctgagc agtatgtctt atcatttgac gcctcccgtc 420

agtctatggg ggccgggtcg cacagcctca cttatgagct cacc 464

<210> 61

<211> 24

<212> DNA

<213> Hepatitis E virus

<220>

<223> us1 733s1

<400> 61

ttgagctcgt tgcaagtcca gacc

24

<210> 62

<211> 22

<212> DNA

<213> Hepatitis E virus

<220>

<223> us2851-r2

<400> 62

ccagagggttg accagggttcg gg

22

<210> 63
 <211> 24
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us1 733s2

<400> 63
 ccgtgctcgg taataagacc ttcc

24

<210> 64
 <211> 433
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us1-432

<400> 64
 ccgtgctcgg taataagacc ttccggacga cgggtggtcga cggcgcccat ctagaggcga 60
 atggccctga gcagtatgtc ttatcatttg acgcctcccg tcagtctatg ggggcccgggt 120
 cgcatagcct cacttatgag ctcacccctg ctggtttgca ggtaggatt tcattctaattg 180
 gtctggattg cactgctaca ttcccccccg gtggagcccc tagcgctgcg cccggggagg 240
 tggcagcctt ttgcagtgcc ctttatagat ataacagggt caccagcgg cactcgctga 300
 ctggcggatt atggttacac cctgaggggt tgcctgggtat tttccccctt ttctccccctg 360
 ggcatatctg ggagtctgcg aaccctttt gcggggaggg gactttgtat acccgaacct 420
 ggtcaacctc tgg 433

<210> 65
 <211> 26
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2851-f1

<400> 65
 gactgtgatt ggtagtcaa tgcctc

26

<210> 66
 <211> 24
 <212> DNA
 <213> Hepatitis E virus

<220>

<223> us1 430-a1

<400> 66

cgtgtcctca gttatggtga gcac

24

<210> 67

<211> 26

<212> DNA

<213> Hepatitis E virus

<220>

<223> us1 430-a2

<400> 67

tattagcctc aaaccaattt gcagcg

26

<210> 68

<211> 382

<212> DNA

<213> Hepatitis E virus

<220>

<223> us1-382

<400> 68

gactgtgatt ggtagtcaa tgccctcaaac ccgggccatc gtcccgaggagg tggcctctgc 60
catgcctttt accaacgttt tccagaagcg ttttaccctaa ctgaattcat catgcgtgag 120
gggtcttgag catcacctt gaccccgcg cctatcattc atgcagtcgc tcccgattat 180
agggttgagc agaaccgaa gaggcttgag gcagcgtacc gtgaaacttg tcccgctcgt 240
ggcaccgctg cctaccgct tttgggttcg ggtatatacc aggtccctgt tagcctcagt 300
tttgatgcct gggaacgtaa tcaccgcccc ggcgatgagc tttacttgac cgagcccgct 360
gcaaattggt ttgaggctaa ta 382

<210> 69

<211> 22

<212> DNA

<213> Hepatitis E virus

<220>

<223> us2-579-s1

<400> 69

cagaccacga gccgtgtgct ac

22

<210> 70

<211> 25

<212> DNA

<213> Hepatitis E virus

<220>

<223> JE hev167-a1

<400> 70

ccaacacact atcggacaca gtgag 25

<210> 71

<211> 22

<212> DNA

<213> Hepatitis E virus

<220>

<223> us2-579-s2

<400> 71

gctgctaagg ctgccaaccc tg 22

<210> 72

<211> 24

<212> DNA

<213> Hepatitis E virus

<220>

<223> JE hev167-a2

<400> 72

cagtgcagctc ctgtggcatg taga 24

<210> 73

<211> 451

<212> DNA

<213> Hepatitis E virus

<220>

<223> us1-579wb

<400> 73

gctgctaagg ctgccaaccc tgggtgcgatt acggtccacg aagctcaggg tgccaccttc 60
acagagacca caatcatagc cacggccgac gccaggggcc ttatccagtc atcccgggct 120
catgctatag ttgcacttac tcgccacact gagaagtgtg ttatcctgga tgcccccggc 180
ctgcttcgtg aggtcggcat ttcggatgtg attgtcaaca actttttcct tgctgggtggc 240
gaggtcggcc rccaccgccc ttctgtgata cctcgcggta accctgatca aaacctcggg 300
actttacagg ccttcccgcc gtccgtgcaa attagtgtt accatcagtt ggctgaggaa 360
ctgggccatc gcccgcccc tgctgcgcc gtcttgcccc cttgccctga gcttgagcag 420
ggcctgctct acatgccaca ggagctcact g 451

<210> 74
<211> 24
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-430s1

<400> 74
ggtatataacc aggtccctgt tagc

24

<210> 75
<211> 22
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-482-a1

<400> 75
ccgctgtgtg aggtgtgaag gc

22

<210> 76
<211> 24
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-430s2

<400> 76
gttagcctca gttttgatgc ctgg

24

<210> 77
<211> 23
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-482-a2

<400> 77
gacgccagct gttacggagc tcc

23

<210> 78
<211> 334
<212> DNA
<213> Hepatitis E virus

<220>
<223> us1-430wb

<400> 78
gttagcctca gttttgatgc ctgggaacgt aatcaccgcc ccggcgatga gctttacttg 60
accgagcccg ctgcaaattg gtttgaggct aataagccgg cgcagccggg gctcaccata 120
actgaggaca cggcccgtag ggccaacctg gcattggaga ttgatgccgc tacagaggtc 180
ggcgtgctt gtgccggttg caccatcagc cctggcattg tgcactatca gtttaccgcc 240
gggggtcccg gctcgggcaa gtcaaggtcc atacaacagg gagatgtcga tgtgggtggtt 300
gtgcccaccc gggagctccg taacagctgg cgtc 334

<210> 79
<211> 23
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-482-s1

<400> 79
gatgtcgatg tgggtggttgc gcc 23

<210> 80
<211> 23
<212> DNA
<213> Hepatitis E virus

<220>
<223> JE us2-579-a1

<400> 80
gtaatcgac cagggttggc agc 23

<210> 81
<211> 23
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-482-s2

<400> 81
ggagctccgt aacagctggc gtc 23

<210> 82
<211> 22
<212> DNA
<213> Hepatitis E virus

<220>

<223> JE us2-579-a2

<400> 82

caggggttggc agccttagca gc

22

<210> 83

<211> 413

<212> DNA

<213> Hepatitis E virus

<220>

<223> us1-482wb

<400> 83

ggagctccgt aacagctggc gtcgccgggg ttttgccggc ttcacacccc acacagcggc 60

ccgtgttact atcggccgcc gcgttgtgat tgatgaggct ccatctctcc cgccacacct 120

gttgctgtta catatgcagc gggcctctc ggtccatctc ctcggtgacc caaatcagat 180

ccctgctatt gatcttgagc acgccggcct ggtccctgcg atccgtcccg agcttgccgc 240

aacgagctgg tggcrcgtta cacaccgttg cccggccgat gtgtgcgagc tcatacgcgg 300

agcctacctt aaaatccaga ccacgagccg tgtgctacgg tcctgtttt ggaatgaacc 360

ggccattggc cagaagttgg ttttcacgca ggctgctaag gctgccaacc ctg 413

<210> 84

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: oligo dT
adapter primer

<400> 84

ggccacgcgt cgactagtac tttttttttt ttttttt

37

<210> 85

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: AUAP primer

<400> 85

ggccacgcgt cgactagtac

20

<210> 86

<211> 22

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer
df-orf3-s1

<400> 86
gcgttggtga ggtgggtcgt gg 22

<210> 87
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer
df-orf3-s2

<400> 87
cgcttcttgg tggtttaccg acag 24

<210> 88
<211> 960
<212> DNA
<213> Hepatitis E virus

<220>
<223> Clone HEV 3p RACE

<400> 88
cgcttcttgg tggtttaccg acagaattga tttcgtcggc tgggggtcaa ctgttttact 60
cccgccctgt tgtctcggcc aatggcgagc caacagtaaa gttatacaca tctgttgaga 120
atgcgcagca agacaagggc atcaccattc cacacgacat agatttaggt gactcccggtg 180
tggttatcca ggattatgat aaccagcacg aacaagatcg acctaccccg tcacctgccc 240
cctcccgccc tttctcagtt cttcgtgcca atgatgtttt gtggctctct ctcactgccg 300
ctgagtacgr ccagaccag tatgggtcgt ccaccaaccc tatgtatgtc tctgatacag 360
tcacgcttgt taatgtagcc actggtgctc aggetgttgc ccgtctctct gactggtcta 420
aagttactct ggatggctgc cctcttacta ccattcagca gtattctaag aaattttatg 480
ttctcccgct tcgsgggaag ctgtcctttt gggaggetgg tacgaccaag gccggctacc 540
cgtataatta taataccact gctagtgacc aaattttgat tgagaacgcg gccggtcacc 600
gtgtcgccat ttctacttat accactagtt tgggtgcgg ccctacctcg atytctgcgg 660
tcggtgtact agctccacat tcggcccttg ctgttctega ggatactgtt gattatcctg 720

ctcgtgccca tacttttgat gatttctgcc cggagtgtcg cacccttggc ctgcagggtt 780
 gtgcattcca atctactatt gctgaacttc agcgtcttaa aatgaaggta ggtaaaaccc 840
 gggagtctta attaattcct tttgtgcccc ctccgcagtt ctctttggct ttattttctca 900
 tttctgcttt ccgcgctncc ctggaaaaaa aaaaaaaaaa gtactagtcg acgcgtggcc 960

<210> 89
 <211> 7202
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us1full

<400> 89
 cctggcatta ctactgcat tgagcaggct gctctggctg cggccaattc tgccttggcg 60
 aatgctgtgg tggttcggcc gtttttatct cgcgtgcaaa ccgagattct tattaatttg 120
 atgcaacccc ggcagttggc tttccgccct gaggtacttt ggaatcacc tatccagcgg 180
 gttatacata atgaattaga acagtactgc cgggctcggg ctggctcgttg cttggagggt 240
 ggagctcacc caagatccat taatgacaac cccaacgttc tgcacgcgtg tttccttaga 300
 ccggttgccc gagatgttca gcgctgggtac tctgccccca ccgcggccc tgcggctaata 360
 tgcgcgcgct ccgcgttgcg tggctctccc ccgcgtgacc gcacttactg ctttgatgga 420
 ttctcccggt gtgcttttgc tgcagagacc ggtgtggctc tttactctct gcatgacct 480
 tggccagctg atgttgacga ggctatggcc cggcacggga tracacgctt gtatgcgcga 540
 ctgcaccttc ccctgaggt gctgctacca ccggcacct accacacaac ctcttatctc 600
 ctgattcacg acggcgaccg cgctgttgta acttacgagg gcgatactag tgcgggctat 660
 aatcatgatg tctccatact tcgtgcgtgg atccgtacta caaaaatagt tggatgatcat 720
 ccgttggtca tagagcgtgt gcgggccatt ggatgtcatt ttgtgttgct gctcaccgca 780
 gcccctgagc cgtcacccat gccttatgtt ccttaccctc gttcaacgga ggtgtatgtc 840
 cggctccatat ttggccctgg cggctcccca tccttgtttc cgtcagcctg ctctactaaa 900
 tctactttcc atgctgtccc ggtgcatatc tgggatcggc tcatgctctt tgggtgccacc 960
 ctggaacgatc aggcgttttg ctgttcacgg ctcatgactt acctccgtgg tattagttac 1020
 aaggtcactg tcggcgcgct tgtcgctaata gaggggtgga acgcctctga agacgctctt 1080
 actgcartga tcaactgcagc ttatttgact atttgccatc agcgttatct ccgcacccag 1140
 gcgatatoca agggcatgcg ccggttgggg gttgagcacg ccagaaaatt tatcacaaga 1200

ctctacagtt ggctatattga gaagtctggc cgtgattata tccccggccg ccagcttcag 1260
 ttctatgcac agtgccgacg gtggctatct gcaggcttcc acctagaccc cagggtactt 1320
 gtttttgatg agtcagtacc atgccgctgt aggacgtttt tgaagaaagt tgcgggtaaa 1380
 ttctgctgtt ttatgcgggtg gctcgggcag gagtgtacct gcttcttgga gccggccgag 1440
 ggttttagtcg gcgatcatgg ccatgacaac gaggcctatg agggttctga ggtcgacccg 1500
 gctgaacctg cacatcttga tgtttctggg acttacgccg tccacgggca ccagcttgag 1560
 gccctctata gggcacttaa tgtcccacaa gatattgccg ctcgagcttc ccgactaacg 1620
 gcaactgttg agctcgttgc aagtccagac cgcttagagt gccgcaccgt gctcggtaat 1680
 aagaccttcc ggacgacggt ggtcgacggc gcccatctag aggcgaaatgg ccctgagcag 1740
 tatgtcttat catttgacgc ctcccgctcag tctatggggg ccgggtcgca tagcctcact 1800
 tatgagctca cccctgctgg tttgcagggt aggatttcat ctaatggtct ggattgcact 1860
 gctacattcc cccccggtgg agccctagc gctgcgcccg gggaggtggc agccttttgc 1920
 agtgccttt atagatataa caggttcacc cagcggcact cgctgactgg cggattatgg 1980
 ttacacctg aggggttget ggggtattttc ccccttttct cccctgggca tatctgggag 2040
 tctgcgaacc ctttttgcgg ggaggggact ttgtataccc gaacttggtc aacatctggc 2100
 ttttctagtg atttctcccc cctgaagcg gccgctcctg ctatggetgc taccgccggg 2160
 ctgccccatt ctacccacc tgttagcgat atttgggtgc taccaccgc ctcagaggag 2220
 tttcagggtg atgcagcacc tgtgccccct gccctgacc ctgctggatt gcccggtccc 2280
 gttgtgctta ccccccccc cctccccct gtgcataagc catcaatacc ccgccttcc 2340
 cgtaaccgtc gtctcctcta tacctatcct gacggcgcta aggtgtatgc agggtcactg 2400
 tttgaatcag actgtgactg gctggttaat gcctcaaacc cgggccatcg tcccggaggt 2460
 ggctctgcc atgcctttta ccaacgtttt ccagaagcgt ttaccccaac tgaattcatc 2520
 atgogtgagg gtcttgacgc atacaccttg accccgcgcc ctatcattca tgcagtcgct 2580
 cccgattata gggttgagca gaaccggaag aggcttgagg cagcgtaacc tgaaacttgt 2640
 tcccgtctg gcaccgctgc ctaccgctt ttgggttcgg gtatatacca ggtccctgtt 2700
 agcctcagtt ttgatgcctg ggaacgtaat caccgccccg gcgatgagct ttacttgacc 2760
 gagcccgctg caaattgggt tgaggctaata agccggcgcc agccggtgct caccataact 2820
 gaggacacgg cccgtacggc caacctggca ttggagattg atgccgctac agaggtcggc 2880

cgtgcttgtg ccggttgcac catcagccct ggcattgtgc actatcagtt taccgccggg 2940
 gtcccgggct cgggcaagtc aaggtccata caacaggag atgtcgatgt ggtggttgtg 3000
 cccacccggg agcttcgtaa tagttggcgc cgccggggtt ttgcggcctt cacaccccaac 3060
 acagcggccc gtgttactat cggccgccgc gttgtgattg atgaggctcc atctctcccg 3120
 ccacacctgt tgctgttaca tatgcagcgg gcctcctcgg tccatctcct cggtgacca 3180
 aatcagatcc ctgctattga ttttgagcac gccggcctgg tccctgcgat ccgtcccgag 3240
 cttgcgcaa cgagctggtg gcrcgttaca caccgttgcc cggccgatgt gtgcgagctc 3300
 atacgaggag cctaccctaa aatccagacc acgagccgtg tgctacggtc cctgttttgg 3360
 aatgaaccgg ccattggcca gaagtgggtt ytcacgcagg cggcaaaggc tgctaaccct 3420
 ggtgcgatta cgggtccacga agctcagggt gccaccttca cagagaccac aatcatagcc 3480
 acggccgacg ccaggggctt tatccagtca tcccgggctc atgctatagt tgcacttact 3540
 cgccacactg agaagtgtgt taccctggat gccccggcc tgcttcgtga ggtcggcatt 3600
 tcggatgtga ttgtcaacaa ctttttcctt gctggtggcg aggtcggccr ccaccgcct 3660
 tctgtgatac ctgcggtaa ccctgatcaa aacctcggga ctttacaggc cttcccgcg 3720
 tcctgtcaaa ttagtgctta ccacagttg gctgaggaaac tgggccatcg ccgggcccct 3780
 gtgcgcgcg tcttgcccc ttgcctgag cttgagcagg gcctgctcta catgccacag 3840
 gagctcactg tgtccgatag tgtgttggtt tttgagctta cggatatagt tcattgccgc 3900
 atggccgctc caagccagcg aaaggctgtt ctctcaacac ttgtggggag gtatggccgt 3960
 aggacgaaac tatatgaggc ggcgattca gatgttcgtg agtccctagc taggttcac 4020
 cctactatcg ggcctgttca ggctaccaca tgtgagttgt atgagttggt tgaggctatg 4080
 gtggagaaag gtcaggacgg ctctgcagtc ttagagcttg atctttgtaa tcgtgatgtc 4140
 tcgcgcacat caatttttcca aaaagwctgc aacaagtta caactggtga gaccatcgcc 4200
 cacggcaagg ttggccaggg tatatcggcc tggagtaaga ccttctgcgc tctgttcggc 4260
 ccgtggttcc gcgccattga aaaagaaata ttggccctgc tccgcctaa tatcttttat 4320
 ggcgacgctt atgaggagtc agtttttgcc gccgtgtgt ccggggcggg gtcatgtatg 4380
 gtatttgaaa atgacttttc agagtttgac agtaccaga ataatttctc tcttggcctt 4440
 gagtgtgtgg ttatggagga gtgcggcatg cctcaatggc taattaggtt gtaccatctg 4500
 gttcggctcg cctggattct gcaggcccg aaggagtctc ttaagggttt ctggaagaag 4560
 cattctggtg agcctgttac cttcttttg aataccgtct ggaatatggc gattatagca 4620

cattgctatg agttccgtga ctttcgtggt gctgccttta agggatgatga ttccgtgggc 4680
ctctgtagtg actaccgaca gagccgcaat gcagctgcct taattgctgg ctgtgggctc 4740
aaattgaagg ttgattaccg ccctatcggg ctgtatgctg gggatgggtg gggcccccgg 4800
ttggggacac tgcccgatgt ggtgcgtttt gctggtcggt tgtctgaaaa gaattggggc 4860
ccggccccgg aacgtgctga gcagctgcgt cttgctgtct gcgacttcct tcgagggttg 4920
acgaatgttg cgcaggctctg tgttgatggt gtgtcccgtg tctatggagt cagccccggg 4980
ctcgtacata acctatttgg catgctgcag accatcgccg atggcaaggc ccactttaca 5040
gagactatta aacctgtact tgatctcaca aattccatca tacagcgggt ggaatgaata 5100
acatgtcttt tgcctcgccc atgggatcac catgcgccct agggctgttc tgttgttgtt 5160
cctcatgttt ctgcctatgc tgcccgcgcc accggccggg cagcgtctg gccgtgcgcg 5220
tgggcggcgc agcggcgggt cggcgggtgg tttctggagt gacagggttg attctcagcc 5280
cttcgcctc ccctatattc atccaaccaa ccccttcgcc gccgatgtcg tttcacaacc 5340
cggggctgga actcgccctc gacagccgcc ccgccccctc ggttcgctt ggctgacca 5400
gtccaagcgc ccttcgttg ccccccgtcg tcgatctacc ccagctgggg ctgcgcgct 5460
aactgccata tcaccagccc ctgatacagc tctgtacct gatgttgact cacgtgggtg 5520
tattttgcgc cggcagtaca atttgtctac gtcccgcctt acatcatctg ttgcttctgg 5580
tactaatctg gttctctatg ctgccccgct gaacctctc ttgcctcttc aggatggcac 5640
caacactcat attatggcta ctgaggcatc taattacgcc cagtatcggg ttgttcgggc 5700
tacgattcgt tatcgcccggt tgggtccaaa tgctgttggg gggttatgcta tctctatttc 5760
tttctggcct caaactacaa ctaccctac ttctgttgac atgaattcta tcacttctac 5820
tgatgtcagg atcttggctc agcccggat agcctccgag ttagtcatcc ctagtgaacg 5880
ccttcactac cgcaaccaag gctggcgctc tgttgagacc acgggtgtgg ccgaagagga 5940
ggctacctcc ggtctggtta tgctttgtat tcatggctcc cctgttaact cctacactaa 6000
tacaccttac accggtgcat tggggcttct tgattttgca ttagaacttg aatttagaaa 6060
tttgacaccc gggaacacta acaccggtgt ttcccggat actagcacag cccgccaccg 6120
gctgcgcgcg ggtgctgatg ggaccgctga gctcaccacc acagcagcca cacgcttcat 6180
gaaggatttg cattttactg gtacgaacgg cgttggtgag gtgggtcgtg gtattgcct 6240
gactctgttt aatcttgctg atacgcttct tgggtggtta ccgacagaat tgatttcgtc 6300

```

ggctgggggt caactgtttt actcccgccc tgttgtctcg gccaatggcg agccaacagt 6360
aaagttatac acatctgttg agaatgcgca gcaagacaag ggcacaccca ttccacacga 6420
catagattta ggtgactccc gtgtggttat ccaggattat gataaccagc acgaacaaga 6480
tcgacctacc ccgtcacctg cccctcccg ccctttctca gttcttcgtg ccaatgatgt 6540
tttgtggctc tctctcactg ccgctgagta cgrccagacc acgtatgggt cgtccaccaa 6600
ccctatgtat gtctctgata cagtcacgct tgttaatgta gccactgggt cttaggctgt 6660
tgcccgctct cttgactggg ctaaagttac tctggatggg cgcctcttta ctaccattca 6720
gcagtattct aagaaatfff atgttctccc gcttcgsggg aagctgtcct tttgggaggc 6780
tggtacgacc aaggccggct acccgataaa ttataatacc actgctagtg accaaatfff 6840
gattgagaac gcggccggct accgtgtcgc catttctact tataccacta gtttgggtgc 6900
cgccctacc tcgatyctg cggtcgggtg actagctcca cattcggccc ttgctgttct 6960
cgaggatact gttgattatc ctgctcgtgc ccatactfff gatgatttct gcccgagtg 7020
tcgcacctt ggtctgcagg gttgtgcatt ccaatctact attgctgaac ttcagcgtct 7080
taaaatgaag gtaggtaaaa cccgggagtc ttaattaatt ctttttgtgc ccccttcgca 7140
gttctctttg gctttatttc tcatttctgc tttccgcgct cctggaaaa aaaaaaaaaa 7200
aa

```

7202

```

<210> 90
<211> 7202
<212> DNA
<213> Hepatitis E virus

```

```

<220>
<221> CDS
<222> (1)..(5097)
<223> orf1

```

```

<220>
<221> CDS
<222> (5132)..(7114)
<223> orf2

```

```

<220>
<221> misc_feature
<222> ( )..( )
<223> CDS- orf3

```

```

<220>
<223> us1full

```

```

<400> 90

```

cct ggc att act act gcc att gag cag gct gct ctg gct gcg gcc aat	48
Pro Gly Ile Thr Thr Ala Ile Glu Gln Ala Ala Leu Ala Ala Ala Asn	
1 5 10 15	
tct gcc ttg gcg aat gct gtg gtg gtt cgg ccg ttt tta tct cgc gtg	96
Ser Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu Ser Arg Val	
20 25 30	
caa acc gag att ctt att aat ttg atg caa ccc cgg cag ttg gtt ttc	144
Gln Thr Glu Ile Leu Ile Asn Leu Met Gln Pro Arg Gln Leu Val Phe	
35 40 45	
cgc cct gag gta ctt tgg aat cac cct atc cag cgg gtt ata cat aat	192
Arg Pro Glu Val Leu Trp Asn His Pro Ile Gln Arg Val Ile His Asn	
50 55 60	
gaa tta gaa cag tac tgc cgg gct cgg gct ggt cgt tgc ttg gag gtt	240
Glu Leu Glu Gln Tyr Cys Arg Ala Arg Ala Gly Arg Cys Leu Glu Val	
65 70 75 80	
gga gct cac cca aga tcc att aat gac aac ccc aac gtt ctg cat cgg	288
Gly Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Leu His Arg	
85 90 95	
tgt ttc ctt aga ccg gtt ggc cga gat gtt cag cgc tgg tac tct gcc	336
Cys Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp Tyr Ser Ala	
100 105 110	
ccc acc cgc ggc cct gcg gct aat tgc cgc cgc tcc gcg ttg cgt ggt	384
Pro Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu Arg Gly	
115 120 125	
ctc ccc ccc gct gac cgc act tac tgc ttt gat gga ttc tcc cgt tgt	432
Leu Pro Pro Ala Asp Arg Thr Tyr Cys Phe Asp Gly Phe Ser Arg Cys	
130 135 140	
gct ttt gct gca gag acc ggt gtg gct ctt tac tct ctg cat gac ctt	480
Ala Phe Ala Ala Glu Thr Gly Val Ala Leu Tyr Ser Leu His Asp Leu	
145 150 155 160	
tgg cca gct gat gtt gca gag gct atg gcc cgc cac ggg atr aca cgc	528
Trp Pro Ala Asp Val Ala Glu Ala Met Ala Arg His Gly Xaa Thr Arg	
165 170 175	
ttg tat gcc gca ctg cac ctt ccc cct gag gtg ctg cta cca ccc ggc	576
Leu Tyr Ala Ala Leu His Leu Pro Pro Glu Val Leu Leu Pro Pro Gly	
180 185 190	
acc tac cac aca acc tcg tat ctc ctg att cac gac ggc gac cgc gct	624
Thr Tyr His Thr Thr Ser Tyr Leu Leu Ile His Asp Gly Asp Arg Ala	
195 200 205	
gtt gta act tac gag ggc gat act agt gcg ggc tat aat cat gat gtc	672
Val Val Thr Tyr Glu Gly Asp Thr Ser Ala Gly Tyr Asn His Asp Val	
210 215 220	
tcc ata ctt cgt gcg tgg atc cgt act aca aaa ata gtt ggt gat cat	720

Ser	Ile	Leu	Arg	Ala	Trp	Ile	Arg	Thr	Thr	Lys	Ile	Val	Gly	Asp	His	
225					230					235					240	
ccg	ttg	gtc	ata	gag	cgt	gtg	cgg	gcc	att	gga	tgt	cat	ttt	gtg	ttg	768
Pro	Leu	Val	Ile	Glu	Arg	Val	Arg	Ala	Ile	Gly	Cys	His	Phe	Val	Leu	
				245					250					255		
ctg	ctc	acc	gca	gcc	cct	gag	cgg	tca	ccc	atg	cct	tat	gtt	cct	tac	816
Leu	Leu	Thr	Ala	Ala	Pro	Glu	Pro	Ser	Pro	Met	Pro	Tyr	Val	Pro	Tyr	
			260					265					270			
cct	cgt	tca	acg	gag	gtg	tat	gtc	cgg	tcc	ata	ttt	ggc	cct	ggc	ggc	864
Pro	Arg	Ser	Thr	Glu	Val	Tyr	Val	Arg	Ser	Ile	Phe	Gly	Pro	Gly	Gly	
		275					280					285				
tcc	cca	tcc	ttg	ttt	ccg	tca	gcc	tgc	tct	act	aaa	tct	act	ttc	cat	912
Ser	Pro	Ser	Leu	Phe	Pro	Ser	Ala	Cys	Ser	Thr	Lys	Ser	Thr	Phe	His	
	290					295					300					
gct	gtc	ccg	gtg	cat	atc	tgg	gat	cgg	ctc	atg	ctc	ttt	ggg	gcc	acc	960
Ala	Val	Pro	Val	His	Ile	Trp	Asp	Arg	Leu	Met	Leu	Phe	Gly	Ala	Thr	
305					310					315					320	
ctg	gac	gat	cag	gcg	ttt	tgc	tgt	tca	cgg	ctc	atg	act	tac	ctc	cgt	1008
Leu	Asp	Asp	Gln	Ala	Phe	Cys	Cys	Ser	Arg	Leu	Met	Thr	Tyr	Leu	Arg	
			325					330						335		
ggg	att	agt	tac	aag	gtc	act	gtc	ggc	gcg	ctt	gtc	gct	aat	gag	ggg	1056
Gly	Ile	Ser	Tyr	Lys	Val	Thr	Val	Gly	Ala	Leu	Val	Ala	Asn	Glu	Gly	
			340					345					350			
tgg	aac	gcc	tct	gaa	gac	gct	ctt	act	gca	rtg	atc	act	gca	gct	tat	1104
Trp	Asn	Ala	Ser	Glu	Asp	Ala	Leu	Thr	Ala	Xaa	Ile	Thr	Ala	Ala	Tyr	
	355						360					365				
ttg	act	att	tgc	cat	cag	cgt	tat	ctc	cgc	acc	cag	gcg	ata	tcc	aag	1152
Leu	Thr	Ile	Cys	His	Gln	Arg	Tyr	Leu	Arg	Thr	Gln	Ala	Ile	Ser	Lys	
	370					375					380					
ggc	atg	cgc	cgg	ttg	ggg	gtt	gag	cac	gcc	cag	aaa	ttt	atc	aca	aga	1200
Gly	Met	Arg	Arg	Leu	Gly	Val	Glu	His	Ala	Gln	Lys	Phe	Ile	Thr	Arg	
385					390					395					400	
ctc	tac	agt	tgg	cta	ttt	gag	aag	tct	ggc	cgt	gat	tat	atc	ccc	ggc	1248
Leu	Tyr	Ser	Trp	Leu	Phe	Glu	Lys	Ser	Gly	Arg	Asp	Tyr	Ile	Pro	Gly	
				405					410					415		
cgc	cag	ctt	cag	ttc	tat	gca	cag	tgc	cga	cgg	tgg	cta	tct	gca	ggc	1296
Arg	Gln	Leu	Gln	Phe	Tyr	Ala	Gln	Cys	Arg	Arg	Trp	Leu	Ser	Ala	Gly	
			420					425					430			
ttc	cac	cta	gac	ccc	agg	gta	ctt	gtt	ttt	gat	gag	tca	gta	cca	tgc	1344
Phe	His	Leu	Asp	Pro	Arg	Val	Leu	Val	Phe	Asp	Glu	Ser	Val	Pro	Cys	
		435					440					445				
cgc	tgt	agg	acg	ttt	ttg	aag	aaa	gtt	gcg	ggg	aaa	ttc	tgc	tgt	ttt	1392
Arg	Cys	Arg	Thr	Phe	Leu	Lys	Lys	Val	Ala	Gly	Lys	Phe	Cys	Cys	Phe	

450	455	460	
atg cgg tgg ctc ggg cag gag tgt acc tgc ttc ttg gag ccg gcc gag			1440
Met Arg Trp Leu Gly Gln Glu Cys Thr Cys Phe Leu Glu Pro Ala Glu			
465	470	475	480
ggt tta gtc ggc gat cat ggc cat gac aac gag gcc tat gag ggt tct			1488
Gly Leu Val Gly Asp His Gly His Asp Asn Glu Ala Tyr Glu Gly Ser			
	485	490	495
gag gtc gac ccg gct gaa cct gca cat ctt gat gtt tct ggg act tac			1536
Glu Val Asp Pro Ala Glu Pro Ala His Leu Asp Val Ser Gly Thr Tyr			
	500	505	510
gcc gtc cac ggg cac cag ctt gag gcc ctc tat agg gca ctt aat gtc			1584
Ala Val His Gly His Gln Leu Glu Ala Leu Tyr Arg Ala Leu Asn Val			
	515	520	525
cca caa gat att gcc gct cga gct tcc cga cta acg gca act gtt gag			1632
Pro Gln Asp Ile Ala Ala Arg Ala Ser Arg Leu Thr Ala Thr Val Glu			
	530	535	540
ctc gtt gca agt cca gac cgc tta gag tgc cgc acc gtg ctc ggt aat			1680
Leu Val Ala Ser Pro Asp Arg Leu Glu Cys Arg Thr Val Leu Gly Asn			
	545	550	560
aag acc ttc ccg acg acg gtg gtc gac ggc gcc cat cta gag gcg aat			1728
Lys Thr Phe Arg Thr Thr Val Val Asp Gly Ala His Leu Glu Ala Asn			
	565	570	575
ggc cct gag cag tat gtc tta tca ttt gac gcc tcc cgt cag tct atg			1776
Gly Pro Glu Gln Tyr Val Leu Ser Phe Asp Ala Ser Arg Gln Ser Met			
	580	585	590
ggg gcc ggg tcg cat agc ctc act tat gag ctc acc cct gct ggt ttg			1824
Gly Ala Gly Ser His Ser Leu Thr Tyr Glu Leu Thr Pro Ala Gly Leu			
	595	600	605
cag gtt agg att tca tct aat ggt ctg gat tgc act gct aca ttc ccc			1872
Gln Val Arg Ile Ser Ser Asn Gly Leu Asp Cys Thr Ala Thr Phe Pro			
	610	615	620
ccc ggt gga gcc cct agc gct gcg ccc ggg gag gtg gca gcc ttt tgc			1920
Pro Gly Gly Ala Pro Ser Ala Ala Pro Gly Glu Val Ala Ala Phe Cys			
	625	630	640
agt gcc ctt tat aga tat aac agg ttc acc cag cgg cac tcg ctg act			1968
Ser Ala Leu Tyr Arg Tyr Asn Arg Phe Thr Gln Arg His Ser Leu Thr			
	645	650	655
ggc gga tta tgg tta cac cct gag ggg ttg ctg ggt att ttc ccc cct			2016
Gly Gly Leu Trp Leu His Pro Glu Gly Leu Leu Gly Ile Phe Pro Pro			
	660	665	670
ttc tcc cct ggg cat atc tgg gag tct gcg aac ccc ttt tgc ggg gag			2064
Phe Ser Pro Gly His Ile Trp Glu Ser Ala Asn Pro Phe Cys Gly Glu			
	675	680	685

ggg Gly 690	act Thr	ttg Leu	tat Tyr	acc Thr	cga Arg	act Thr	tgg Trp	tca Ser	aca Thr	tct Ser	ggc Gly	ttt Phe	tct Ser	agt Ser	gat Asp	2112
ttc Phe 705	tcc Ser	ccc Pro	cct Pro	gaa Glu	gcg Ala	gcc Ala	gct Ala	cct Pro	gct Ala	atg Met	gct Ala	gct Ala	acc Thr	ccg Pro	ggg Gly	2160
ctg Leu	ccc Pro	cat His	tct Ser	acc Thr	cca Pro	cct Pro	gtt Val	agc Ser	gat Asp	att Ile	tgg Trp	gtg Val	cta Leu	cca Pro	ccg Pro	2208
ccc Pro	tca Ser	gag Glu	gag Glu	ttt Phe	cag Gln	gtt Val	gat Asp	gca Ala	gca Ala	cct Pro	gtg Val	ccc Pro	cct Pro	gcc Ala	cct Pro	2256
gac Asp	cct Pro	gct Ala	gga Gly	ttg Leu	ccc Pro	ggg Gly	ccc Pro	gtt Val	gtg Val	ctt Leu	acc Thr	ccc Pro	ccc Pro	ccc Pro	cct Pro	2304
ccc Pro	cct Pro	gtg Val	cat His	aag Lys	cca Pro	tca Ser	ata Ile	ccc Pro	ccg Pro	cct Pro	tcc Ser	cgt Arg	aac Asn	cgt Arg	cgt Arg	2352
ctc Leu	ctc Leu	tat Tyr	acc Thr	tat Tyr	cct Pro	gac Asp	ggc Gly	gct Ala	aag Lys	gtg Val	tat Tyr	gca Ala	ggg Gly	tca Ser	ctg Leu	2400
ttt Phe	gaa Glu	tca Ser	gac Asp	tgt Cys	gac Asp	tgg Trp	ctg Leu	gtt Val	aat Asn	gcc Ala	tca Ser	aac Asn	ccg Pro	ggc Gly	cat His	2448
cgt Arg	ccc Pro	gga Gly	ggg Gly	ggc Gly	ctc Leu	tgc Cys	cat His	gcc Ala	ttt Phe	tac Tyr	caa Gln	cgt Arg	ttt Phe	cca Pro	gaa Glu	2496
gcg Ala	ttt Phe	tac Tyr	cca Pro	act Thr	gaa Glu	ttc Phe	atc Ile	atg Met	cgt Arg	gag Glu	ggg Gly	ctt Leu	gca Ala	gca Ala	tac Tyr	2544
acc Thr	ttg Leu	acc Thr	ccg Pro	cgc Arg	cct Pro	atc Ile	att Ile	cat His	gca Ala	gtc Val	gct Ala	ccc Pro	gat Asp	tat Tyr	agg Arg	2592
gtt Val	gag Glu	cag Gln	aac Asn	ccg Pro	aag Lys	agg Arg	ctt Leu	gag Glu	gca Ala	gcg Ala	tac Tyr	cgt Arg	gaa Glu	act Thr	tgt Cys	2640
tcc Ser	cgt Arg	cgt Arg	ggc Gly	acc Thr	gct Ala	gcc Ala	tac Tyr	ccg Pro	ctt Leu	ttg Leu	ggg Gly	tgc Ser	ggg Gly	ata Ile	tac Tyr	2688
cag Gln	gtc Val	cct Pro	gtt Val	agc Ser	ctc Leu	agt Ser	ttt Phe	gat Asp	gcc Ala	tgg Trp	gaa Glu	cgt Arg	aat Asn	cac His	cgc Arg	2736

ccc ggc gat gag ctt tac ttg acc gag ccc gct gca aat tgg ttt gag	2784
Pro Gly Asp Glu Leu Tyr Leu Thr Glu Pro Ala Ala Asn Trp Phe Glu	
915 920 925	
gct aat aag ccg gcg cag ccg gtg ctc acc ata act gag gac acg gcc	2832
Ala Asn Lys Pro Ala Gln Pro Val Leu Thr Ile Thr Glu Asp Thr Ala	
930 935 940	
cgt acg gcc aac ctg gca ttg gag att gat gcc gct aca gag gtc ggc	2880
Arg Thr Ala Asn Leu Ala Leu Glu Ile Asp Ala Ala Thr Glu Val Gly	
945 950 955 960	
cgt gct tgt gcc ggt tgc acc atc agc cct ggc att gtg cac tat cag	2928
Arg Ala Cys Ala Gly Cys Thr Ile Ser Pro Gly Ile Val His Tyr Gln	
965 970 975	
ttt acc gcc ggg gtc ccg ggc tgc ggc aag tca agg tcc ata caa cag	2976
Phe Thr Ala Gly Val Pro Gly Ser Gly Lys Ser Arg Ser Ile Gln Gln	
980 985 990	
gga gat gtc gat gtg gtg gtt gtg ccc acc cgg gag ctt cgt aat agt	3024
Gly Asp Val Asp Val Val Val Pro Thr Arg Glu Leu Arg Asn Ser	
995 1000 1005	
tgg cgc cgc cgg ggt ttt gcg gcc ttc aca ccc cac aca gcg gcc cgt	3072
Trp Arg Arg Arg Gly Phe Ala Ala Phe Thr Pro His Thr Ala Ala Arg	
1010 1015 1020	
gtt act atc gcc cgc cgc gtt gtg att gat gag gct cca tct ctc ccg	3120
Val Thr Ile Gly Arg Arg Val Val Ile Asp Glu Ala Pro Ser Leu Pro	
1025 1030 1035 1040	
cca cac ctg ttg ctg tta cat atg cag cgg gcc tcc tgc gtc cat ctc	3168
Pro His Leu Leu Leu Leu His Met Gln Arg Ala Ser Ser Val His Leu	
1045 1050 1055	
ctc ggt gac cca aat cag atc cct gct att gat ttt gag cac gcc ggc	3216
Leu Gly Asp Pro Asn Gln Ile Pro Ala Ile Asp Phe Glu His Ala Gly	
1060 1065 1070	
ctg gtc cct gcg atc cgt ccc gag ctt gcg cca acg agc tgg tgg crc	3264
Leu Val Pro Ala Ile Arg Pro Glu Leu Ala Pro Thr Ser Trp Trp Xaa	
1075 1080 1085	
gtt aca cac cgt tgc ccg gcc gat gtg tgc gag ctc ata cgc gga gcc	3312
Val Thr His Arg Cys Pro Ala Asp Val Cys Glu Leu Ile Arg Gly Ala	
1090 1095 1100	
tac cct aaa atc cag acc acg agc cgt gtg cta cgg tcc ctg ttt tgg	3360
Tyr Pro Lys Ile Gln Thr Thr Ser Arg Val Leu Arg Ser Leu Phe Trp	
1105 1110 1115 1120	
aat gaa ccg gcc att ggc cag aag ttg gtt ytc acg cag gcg gca aag	3408
Asn Glu Pro Ala Ile Gly Gln Lys Leu Val Xaa Thr Gln Ala Ala Lys	
1125 1130 1135	
gct gct aac cct ggt gcg att acg gtc cac gaa gct cag ggt gcc acc	3456

Ala Ala Asn Pro Gly Ala Ile Thr Val His Glu Ala Gln Gly Ala Thr	
1140 1145 1150	
ttc aca gag acc aca atc ata gcc acg gcc gac gcc agg ggc ctt atc	3504
Phe Thr Glu Thr Thr Ile Ile Ala Thr Ala Asp Ala Arg Gly Leu Ile	
1155 1160 1165	
cag tca tcc cgg gct cat gct ata gtt gca ctt act cgc cac act gag	3552
Gln Ser Ser Arg Ala His Ala Ile Val Ala Leu Thr Arg His Thr Glu	
1170 1175 1180	
aag tgt gtt atc ctg gat gcc ccc ggc ctg ctt cgt gag gtc ggc att	3600
Lys Cys Val Ile Leu Asp Ala Pro Gly Leu Leu Arg Glu Val Gly Ile	
1185 1190 1195 1200	
tcg gat gtg att gtc aac aac ttt ttc ctt gct ggt ggc gag gtc ggc	3648
Ser Asp Val Ile Val Asn Asn Phe Phe Leu Ala Gly Gly Glu Val Gly	
1205 1210 1215	
crc cac cgc cct tct gtg ata cct cgc ggt aac cct gat caa aac ctc	3696
Xaa His Arg Pro Ser Val Ile Pro Arg Gly Asn Pro Asp Gln Asn Leu	
1220 1225 1230	
ggg act tta cag gcc ttc ccg ccg tcc tgt caa att agt gct tac cat	3744
Gly Thr Leu Gln Ala Phe Pro Pro Ser Cys Gln Ile Ser Ala Tyr His	
1235 1240 1245	
cag ttg gct gag gaa ctg ggc cat cgc ccg gcc cct gtc gcc gcc gtc	3792
Gln Leu Ala Glu Glu Leu Gly His Arg Pro Ala Pro Val Ala Ala Val	
1250 1255 1260	
ttg ccc cct tgc cct gag ctt gag cag ggc ctg ctc tac atg cca cag	3840
Leu Pro Pro Cys Pro Glu Leu Glu Gln Gly Leu Leu Tyr Met Pro Gln	
1265 1270 1275 1280	
gag ctc act gtg tcc gat agt gtg ttg gtt ttt gag ctt acg gat ata	3888
Glu Leu Thr Val Ser Asp Ser Val Leu Val Phe Glu Leu Thr Asp Ile	
1285 1290 1295	
gtt cat tgc cgc atg gcc gct cca agc cag cga aag gct gtt ctc tca	3936
Val His Cys Arg Met Ala Ala Pro Ser Gln Arg Lys Ala Val Leu Ser	
1300 1305 1310	
aca ctt gtg ggg agg tat ggc cgt agg acg aaa cta tat gag gcg gcg	3984
Thr Leu Val Gly Arg Tyr Gly Arg Arg Thr Lys Leu Tyr Glu Ala Ala	
1315 1320 1325	
cat tca gat gtt cgt gag tcc cta gct agg ttc atc cct act atc ggg	4032
His Ser Asp Val Arg Glu Ser Leu Ala Arg Phe Ile Pro Thr Ile Gly	
1330 1335 1340	
cct gtt cag gct acc aca tgt gag ttg tat gag ttg gtt gag gct atg	4080
Pro Val Gln Ala Thr Thr Cys Glu Leu Tyr Glu Leu Val Glu Ala Met	
1345 1350 1355 1360	
gtg gag aaa ggt cag gac ggc tct gca gtc tta gag ctt gat ctt tgt	4128
Val Glu Lys Gly Gln Asp Gly Ser Ala Val Leu Glu Leu Asp Leu Cys	

1365	1370	1375	
aat cgt gat gtc tcg cgc atc aca ttt ttc caa aaa gwc tgc aac aag			4176
Asn Arg Asp Val Ser Arg Ile Thr Phe Phe Gln Lys Xaa Cys Asn Lys			
1380	1385	1390	
ttt aca act ggt gag acc atc gcc cac ggc aag gtt ggc cag ggt ata			4224
Phe Thr Thr Gly Glu Thr Ile Ala His Gly Lys Val Gly Gln Gly Ile			
1395	1400	1405	
tcg gcc tgg agt aag acc ttc tgc gct ctg ttc ggc ccg tgg ttc cgc			4272
Ser Ala Trp Ser Lys Thr Phe Cys Ala Leu Phe Gly Pro Trp Phe Arg			
1410	1415	1420	
gcc att gaa aaa gaa ata ttg gcc ctg ctc ccg cct aat atc ttt tat			4320
Ala Ile Glu Lys Glu Ile Leu Ala Leu Leu Pro Pro Asn Ile Phe Tyr			
1425	1430	1435	1440
ggc gac gct tat gag gag tca gtt ttt gcc gcc gct gtg tcc ggg gcg			4368
Gly Asp Ala Tyr Glu Glu Ser Val Phe Ala Ala Ala Val Ser Gly Ala			
1445	1450	1455	
ggg tca tgt atg gta ttt gaa aat gac ttt tca gag ttt gac agt acc			4416
Gly Ser Cys Met Val Phe Glu Asn Asp Phe Ser Glu Phe Asp Ser Thr			
1460	1465	1470	
cag aat aat ttc tct ctt ggc ctt gag tgt gtg gtt atg gag gag tgc			4464
Gln Asn Asn Phe Ser Leu Gly Leu Glu Cys Val Val Met Glu Glu Cys			
1475	1480	1485	
ggc atg cct caa tgg cta att agg ttg tac cat ctg gtt ccg tct gcc			4512
Gly Met Pro Gln Trp Leu Ile Arg Leu Tyr His Leu Val Arg Ser Ala			
1490	1495	1500	
tgg att ctg cag gcg ccg aag gag tct ctt aag ggt ttc tgg aag aag			4560
Trp Ile Leu Gln Ala Pro Lys Glu Ser Leu Lys Gly Phe Trp Lys Lys			
1505	1510	1515	1520
cat tct ggt gag cct ggt acc ctt ctt tgg aat acc gtc tgg aat atg			4608
His Ser Gly Glu Pro Gly Thr Leu Leu Trp Asn Thr Val Trp Asn Met			
1525	1530	1535	
gcg att ata gca cat tgc tat gag ttc cgt gac ttt cgt gtt gct gcc			4656
Ala Ile Ile Ala His Cys Tyr Glu Phe Arg Asp Phe Arg Val Ala Ala			
1540	1545	1550	
ttt aag ggt gat gat tcg gtg gtc ctc tgt agt gac tac cga cag agc			4704
Phe Lys Gly Asp Asp Ser Val Val Leu Cys Ser Asp Tyr Arg Gln Ser			
1555	1560	1565	
cgc aat gca gct gcc tta att gct ggc tgt ggg ctc aaa ttg aag gtt			4752
Arg Asn Ala Ala Ala Leu Ile Ala Gly Cys Gly Leu Lys Leu Lys Val			
1570	1575	1580	
gat tac cgc cct atc ggg ctg tat gct ggg gtg gtg gtg gcc ccc ggt			4800
Asp Tyr Arg Pro Ile Gly Leu Tyr Ala Gly Val Val Val Ala Pro Gly			
1585	1590	1595	1600

ttg ggg aca ctg ccc gat gtg gtg cgt ttt gct ggt cgg ttg tct gaa	4848
Leu Gly Thr Leu Pro Asp Val Val Arg Phe Ala Gly Arg Leu Ser Glu	
1605 1610 1615	
aag aat tgg ggc ccc ggc ccg gaa cgt gct gag cag ctg cgt ctt gct	4896
Lys Asn Trp Gly Pro Gly Pro Glu Arg Ala Glu Gln Leu Arg Leu Ala	
1620 1625 1630	
gtc tgc gac ttc ctt cga ggg ttg acg aat gtt gcg cag gtc tgt gtt	4944
Val Cys Asp Phe Leu Arg Gly Leu Thr Asn Val Ala Gln Val Cys Val	
1635 1640 1645	
gat gtt gtg tcc cgt gtc tat gga gtc agc ccc ggg ctc gta cat aac	4992
Asp Val Val Ser Arg Val Tyr Gly Val Ser Pro Gly Leu Val His Asn	
1650 1655 1660	
ctt att ggc atg ctg cag acc atc gcc gat ggc aag gcc cac ttt aca	5040
Leu Ile Gly Met Leu Gln Thr Ile Ala Asp Gly Lys Ala His Phe Thr	
1665 1670 1675 1680	
gag act att aaa cct gta ctt gat ctc aca aat tcc atc ata cag cgg	5088
Glu Thr Ile Lys Pro Val Leu Asp Leu Thr Asn Ser Ile Ile Gln Arg	
1685 1690 1695	
gtg gaa tga ataacatgtc ttttgcacgc cccatgggat cacc atg cgc cct agg	5143
Val Glu Met Arg Pro Arg	
1700	
gct gtt ctg ttg ttg ttc ctc atg ttt ctg cct atg ctg ccc gcg cca	5191
Ala Val Leu Leu Leu Phe Leu Met Phe Leu Pro Met Leu Pro Ala Pro	
1705 1710 1715	
ccg gcc ggt cag ccg tct ggc cgt cgc cgt ggg cgg cgc agc ggc ggt	5239
Pro Ala Gly Gln Pro Ser Gly Arg Arg Arg Gly Arg Arg Ser Gly Gly	
1720 1725 1730 1735	
gcc gcc ggt ggt ttc tgg agt gac agg gtt gat tct cag ccc ttc gcc	5287
Ala Gly Gly Gly Phe Trp Ser Asp Arg Val Asp Ser Gln Pro Phe Ala	
1740 1745 1750	
ctc ccc tat att cat cca acc aac ccc ttc gcc gcc gat gtc gtt tca	5335
Leu Pro Tyr Ile His Pro Thr Asn Pro Phe Ala Ala Asp Val Val Ser	
1755 1760 1765	
caa ccc ggg gct gga act cgc cct cga cag ccg ccc cgc ccc ctc ggt	5383
Gln Pro Gly Ala Gly Thr Arg Pro Arg Gln Pro Pro Arg Pro Leu Gly	
1770 1775 1780	
tcc gct tgg cgt gac cag tcc aag cgc ccc tcc gtt gcc ccc cgt cgt	5431
Ser Ala Trp Arg Asp Gln Ser Lys Arg Pro Ser Val Ala Pro Arg Arg	
1785 1790 1795	
cga tct acc cca gct ggg gct gcg ccg cta act gcc ata tca cca gcc	5479
Arg Ser Thr Pro Ala Gly Ala Ala Pro Leu Thr Ala Ile Ser Pro Ala	
1800 1805 1810 1815	

cct gat aca gct cct gta cct gat gtt gac tca cgt ggt gct att ttg	5527
Pro Asp Thr Ala Pro Val Pro Asp Val Asp Ser Arg Gly Ala Ile Leu	
1820 1825 1830	
cgc cgg cag tac aat ttg tct acg tcc ccg ctt aca tca tct gtt gct	5575
Arg Arg Gln Tyr Asn Leu Ser Thr Ser Pro Leu Thr Ser Ser Val Ala	
1835 1840 1845	
tct ggt act aat ctg gtt ctc tat gct gcc ccg ctg aac cct ctc ttg	5623
Ser Gly Thr Asn Leu Val Leu Tyr Ala Ala Pro Leu Asn Pro Leu Leu	
1850 1855 1860	
cct ctt cag gat ggc acc aac act cat att atg gct act gag gca tct	5671
Pro Leu Gln Asp Gly Thr Asn Thr His Ile Met Ala Thr Glu Ala Ser	
1865 1870 1875	
aat tac gcc cag tat cgg gtt gtt cgg gct acg att cgt tat cgc ccg	5719
Asn Tyr Ala Gln Tyr Arg Val Val Arg Ala Thr Ile Arg Tyr Arg Pro	
1880 1885 1890 1895	
ttg gtg cca aat gct gtt ggt ggt tat gct atc tct att tct ttc tgg	5767
Leu Val Pro Asn Ala Val Gly Gly Tyr Ala Ile Ser Ile Ser Phe Trp	
1900 1905 1910	
cct caa act aca act acc cct act tct gtt gac atg aat tct atc act	5815
Pro Gln Thr Thr Thr Pro Thr Val Asp Met Asn Ser Ile Thr	
1915 1920 1925	
tct act gat gtc agg atc ttg gtc cag ccc ggt ata gcc tcc gag tta	5863
Ser Thr Asp Val Arg Ile Leu Val Gln Pro Gly Ile Ala Ser Glu Leu	
1930 1935 1940	
gtc atc cct agt gaa cgc ctt cac tac cgc aac caa ggc tgg cgc tct	5911
Val Ile Pro Ser Glu Arg Leu His Tyr Arg Asn Gln Gly Trp Arg Ser	
1945 1950 1955	
gtt gag acc acg ggt gtg gcc gaa gag gag gct acc tcc ggt ctg gta	5959
Val Glu Thr Thr Gly Val Ala Glu Glu Glu Ala Thr Ser Gly Leu Val	
1960 1965 1970 1975	
atg ctt tgt att cat ggc tcc cct gtt aac tcc tac act aat aca cct	6007
Met Leu Cys Ile His Gly Ser Pro Val Asn Ser Tyr Thr Asn Thr Pro	
1980 1985 1990	
tac acc ggt gca ttg ggg ctt ctt gat ttt gca tta gaa ctt gaa ttt	6055
Tyr Thr Gly Ala Leu Gly Leu Leu Asp Phe Ala Leu Glu Leu Glu Phe	
1995 2000 2005	
aga aat ttg aca ccc ggg aac act aac acc cgt gtt tcc cgg tat act	6103
Arg Asn Leu Thr Pro Gly Asn Thr Asn Thr Arg Val Ser Arg Tyr Thr	
2010 2015 2020	
agc aca gcc cgc cac cgg ctg cgc cgc ggt gct gat ggg acc gct gag	6151
Ser Thr Ala Arg His Arg Leu Arg Arg Gly Ala Asp Gly Thr Ala Glu	
2025 2030 2035	
ctc acc acc aca gca gcc aca cgc ttc atg aag gat ttg cat ttt act	6199

Leu	Thr	Thr	Thr	Ala	Ala	Thr	Arg	Phe	Met	Lys	Asp	Leu	His	Phe	Thr	
2040					2045					2050					2055	
ggt	acg	aac	ggc	gtt	ggt	gag	gtg	ggt	cgt	ggt	att	gcc	ctg	act	ctg	6247
Gly	Thr	Asn	Gly	Val	Gly	Glu	Val	Gly	Arg	Gly	Ile	Ala	Leu	Thr	Leu	
				2060					2065					2070		
ttt	aat	ctt	gct	gat	acg	ctt	ctt	ggt	ggt	tta	ccg	aca	gaa	ttg	att	6295
Phe	Asn	Leu	Ala	Asp	Thr	Leu	Leu	Gly	Gly	Leu	Pro	Thr	Glu	Leu	Ile	
			2075					2080					2085			
tcg	tcg	gct	ggg	ggt	caa	ctg	ttt	tac	tcc	cgc	cct	gtt	gtc	tcg	gcc	6343
Ser	Ser	Ala	Gly	Gly	Gln	Leu	Phe	Tyr	Ser	Arg	Pro	Val	Val	Ser	Ala	
		2090					2095					2100				
aat	ggc	gag	cca	aca	gta	aag	tta	tac	aca	tct	gtt	gag	aat	gcg	cag	6391
Asn	Gly	Glu	Pro	Thr	Val	Lys	Leu	Tyr	Thr	Ser	Val	Glu	Asn	Ala	Gln	
	2105					2110					2115					
caa	gac	aag	ggc	atc	acc	att	cca	cac	gac	ata	gat	tta	ggg	gac	tcc	6439
Gln	Asp	Lys	Gly	Ile	Thr	Ile	Pro	His	Asp	Ile	Asp	Leu	Gly	Asp	Ser	
2120				2125					2130					2135		
cgt	gtg	gtt	atc	cag	gat	tat	gat	aac	cag	cac	gaa	caa	gat	cga	cct	6487
Arg	Val	Val	Ile	Gln	Asp	Tyr	Asp	Asn	Gln	His	Glu	Gln	Asp	Arg	Pro	
			2140					2145					2150			
acc	ccg	tca	cct	gcc	ccc	tcc	cgc	cct	ttc	tca	gtt	ctt	cgt	gcc	aat	6535
Thr	Pro	Ser	Pro	Ala	Pro	Ser	Arg	Pro	Phe	Ser	Val	Leu	Arg	Ala	Asn	
		2155						2160				2165				
gat	gtt	ttg	tgg	ctc	tct	ctc	act	gcc	gct	gag	tac	grc	cag	acc	acg	6583
Asp	Val	Leu	Trp	Leu	Ser	Leu	Thr	Ala	Ala	Glu	Tyr	Xaa	Gln	Thr	Thr	
	2170						2175				2180					
tat	ggg	tcg	tcc	acc	aac	cct	atg	tat	gtc	tct	gat	aca	gtc	acg	ctt	6631
Tyr	Gly	Ser	Ser	Thr	Asn	Pro	Met	Tyr	Val	Ser	Asp	Thr	Val	Thr	Leu	
	2185					2190					2195					
gtt	aat	gta	gcc	act	ggt	gct	cag	gct	gtt	gcc	cgc	tct	ctt	gac	tgg	6679
Val	Asn	Val	Ala	Thr	Gly	Ala	Gln	Ala	Val	Ala	Arg	Ser	Leu	Asp	Trp	
2200				2205					2210				2215			
tct	aaa	gtt	act	ctg	gat	ggt	cgc	cct	ctt	act	acc	att	cag	cag	tat	6727
Ser	Lys	Val	Thr	Leu	Asp	Gly	Arg	Pro	Leu	Thr	Thr	Ile	Gln	Gln	Tyr	
			2220					2225				2230				
tct	aag	aaa	ttt	tat	gtt	ctc	ccg	ctt	cgs	ggg	aag	ctg	tcc	ttt	tgg	6775
Ser	Lys	Lys	Phe	Tyr	Val	Leu	Pro	Leu	Xaa	Gly	Lys	Leu	Ser	Phe	Trp	
	2235						2240				2245					
gag	gct	ggt	acg	acc	aag	gcc	ggc	tac	ccg	tat	aat	tat	aat	acc	act	6823
Glu	Ala	Gly	Thr													

2265 2270 2275
 att tct act tat acc act agt ttg ggt gcc ggc cct acc tcg aty tct 6919
 Ile Ser Thr Tyr Thr Thr Ser Leu Gly Ala Gly Pro Thr Ser Xaa Ser
 2280 2285 2290 2295

 gcg gtc ggt gta cta gct cca cat tcg gcc ctt gct gtt ctc gag gat 6967
 Ala Val Gly Val Leu Ala Pro His Ser Ala Leu Ala Val Leu Glu Asp
 2300 2305 2310

 act gtt gat tat cct gct cgt gcc cat act ttt gat gat ttc tgc ccg 7015
 Thr Val Asp Tyr Pro Ala Arg Ala His Thr Phe Asp Asp Phe Cys Pro
 2315 2320 2325

 gag tgt cgc acc ctt ggt ctg cag ggt tgt gca ttc caa tct act att 7063
 Glu Cys Arg Thr Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr Ile
 2330 2335 2340

 gct gaa ctt cag cgt ctt aaa atg aag gta ggt aaa acc cgg gag tct 7111
 Ala Glu Leu Gln Arg Leu Lys Met Lys Val Gly Lys Thr Arg Glu Ser
 2345 2350 2355

 taa ttaattcctt ttgtgcccc ttcgcagttc tctttggctt tattttctcat 7164
 2360

 ttctgctttc cgcgctccct ggaaaaaaaa aaaaaaaaa 7202

 <210> 91
 <211> 1698
 <212> PRT
 <213> Hepatitis E virus

 <400> 91
 Pro Gly Ile Thr Thr Ala Ile Glu Gln Ala Ala Leu Ala Ala Ala Asn
 1 5 10 15

 Ser Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu Ser Arg Val
 20 25 30

 Gln Thr Glu Ile Leu Ile Asn Leu Met Gln Pro Arg Gln Leu Val Phe
 35 40 45

 Arg Pro Glu Val Leu Trp Asn His Pro Ile Gln Arg Val Ile His Asn
 50 55 60

 Glu Leu Glu Gln Tyr Cys Arg Ala Arg Ala Gly Arg Cys Leu Glu Val
 65 70 75 80

 Gly Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Leu His Arg
 85 90 95

 Cys Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp Tyr Ser Ala
 100 105 110

 Pro Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu Arg Gly

115					120					125					
Leu	Pro	Pro	Ala	Asp	Arg	Thr	Tyr	Cys	Phe	Asp	Gly	Phe	Ser	Arg	Cys
130						135					140				
Ala	Phe	Ala	Ala	Glu	Thr	Gly	Val	Ala	Leu	Tyr	Ser	Leu	His	Asp	Leu
145					150					155					160
Trp	Pro	Ala	Asp	Val	Ala	Glu	Ala	Met	Ala	Arg	His	Gly	Xaa	Thr	Arg
				165					170					175	
Leu	Tyr	Ala	Ala	Leu	His	Leu	Pro	Pro	Glu	Val	Leu	Leu	Pro	Pro	Gly
			180					185					190		
Thr	Tyr	His	Thr	Thr	Ser	Tyr	Leu	Leu	Ile	His	Asp	Gly	Asp	Arg	Ala
		195					200					205			
Val	Val	Thr	Tyr	Glu	Gly	Asp	Thr	Ser	Ala	Gly	Tyr	Asn	His	Asp	Val
	210					215					220				
Ser	Ile	Leu	Arg	Ala	Trp	Ile	Arg	Thr	Thr	Lys	Ile	Val	Gly	Asp	His
225					230					235					240
Pro	Leu	Val	Ile	Glu	Arg	Val	Arg	Ala	Ile	Gly	Cys	His	Phe	Val	Leu
				245					250					255	
Leu	Leu	Thr	Ala	Ala	Pro	Glu	Pro	Ser	Pro	Met	Pro	Tyr	Val	Pro	Tyr
			260					265					270		
Pro	Arg	Ser	Thr	Glu	Val	Tyr	Val	Arg	Ser	Ile	Phe	Gly	Pro	Gly	Gly
		275					280					285			
Ser	Pro	Ser	Leu	Phe	Pro	Ser	Ala	Cys	Ser	Thr	Lys	Ser	Thr	Phe	His
	290					295					300				
Ala	Val	Pro	Val	His	Ile	Trp	Asp	Arg	Leu	Met	Leu	Phe	Gly	Ala	Thr
305					310					315					320
Leu	Asp	Asp	Gln	Ala	Phe	Cys	Cys	Ser	Arg	Leu	Met	Thr	Tyr	Leu	Arg
				325					330					335	
Gly	Ile	Ser	Tyr	Lys	Val	Thr	Val	Gly	Ala	Leu	Val	Ala	Asn	Glu	Gly
			340					345					350		
Trp	Asn	Ala	Ser	Glu	Asp	Ala	Leu	Thr	Ala	Xaa	Ile	Thr	Ala	Ala	Tyr
		355					360					365			
Leu	Thr	Ile	Cys	His	Gln	Arg	Tyr	Leu	Arg	Thr	Gln	Ala	Ile	Ser	Lys
	370					375					380				
Gly	Met	Arg	Arg	Leu	Gly	Val	Glu	His	Ala	Gln	Lys	Phe	Ile	Thr	Arg
385					390					395					400
Leu	Tyr	Ser	Trp	Leu	Phe	Glu	Lys	Ser	Gly	Arg	Asp	Tyr	Ile	Pro	Gly
				405					410					415	
Arg	Gln	Leu	Gln	Phe	Tyr	Ala	Gln	Cys	Arg	Arg	Trp	Leu	Ser	Ala	Gly

420					425					430						
Phe	His	Leu	Asp	Pro	Arg	Val	Leu	Val	Phe	Asp	Glu	Ser	Val	Pro	Cys	
435					440					445						
Arg	Cys	Arg	Thr	Phe	Leu	Lys	Lys	Val	Ala	Gly	Lys	Phe	Cys	Cys	Phe	
450					455					460						
Met	Arg	Trp	Leu	Gly	Gln	Glu	Cys	Thr	Cys	Phe	Leu	Glu	Pro	Ala	Glu	
465					470					475					480	
Gly	Leu	Val	Gly	Asp	His	Gly	His	Asp	Asn	Glu	Ala	Tyr	Glu	Gly	Ser	
485					490					495						
Glu	Val	Asp	Pro	Ala	Glu	Pro	Ala	His	Leu	Asp	Val	Ser	Gly	Thr	Tyr	
500					505					510						
Ala	Val	His	Gly	His	Gln	Leu	Glu	Ala	Leu	Tyr	Arg	Ala	Leu	Asn	Val	
515					520					525						
Pro	Gln	Asp	Ile	Ala	Ala	Arg	Ala	Ser	Arg	Leu	Thr	Ala	Thr	Val	Glu	
530					535					540						
Leu	Val	Ala	Ser	Pro	Asp	Arg	Leu	Glu	Cys	Arg	Thr	Val	Leu	Gly	Asn	
545					550					555					560	
Lys	Thr	Phe	Arg	Thr	Thr	Val	Val	Asp	Gly	Ala	His	Leu	Glu	Ala	Asn	
565					570					575						
Gly	Pro	Glu	Gln	Tyr	Val	Leu	Ser	Phe	Asp	Ala	Ser	Arg	Gln	Ser	Met	
580					585					590						
Gly	Ala	Gly	Ser	His	Ser	Leu	Thr	Tyr	Glu	Leu	Thr	Pro	Ala	Gly	Leu	
595					600					605						
Gln	Val	Arg	Ile	Ser	Ser	Asn	Gly	Leu	Asp	Cys	Thr	Ala	Thr	Phe	Pro	
610					615					620						
Pro	Gly	Gly	Ala	Pro	Ser	Ala	Ala	Pro	Gly	Glu	Val	Ala	Ala	Phe	Cys	
625					630					635					640	
Ser	Ala	Leu	Tyr	Arg	Tyr	Asn	Arg	Phe	Thr	Gln	Arg	His	Ser	Leu	Thr	
645					650					655						
Gly	Gly	Leu	Trp	Leu	His	Pro	Glu	Gly	Leu	Leu	Gly	Ile	Phe	Pro	Pro	
660					665					670						
Phe	Ser	Pro	Gly	His	Ile	Trp	Glu	Ser	Ala	Asn	Pro	Phe	Cys	Gly	Glu	
675					680					685						
Gly	Thr	Leu	Tyr	Thr	Arg	Thr	Trp	Ser	Thr	Ser	Gly	Phe	Ser	Ser	Asp	
690					695					700						
Phe	Ser	Pro	Pro	Glu	Ala	Ala	Ala	Pro	Ala	Met	Ala	Ala	Thr	Pro	Gly	
705					710					715					720	
Leu	Pro	His	Ser	Thr	Pro	Pro	Val	Ser	Asp	Ile	Trp	Val	Leu	Pro	Pro	

725										730					735				
Pro	Ser	Glu	Glu	Phe	Gln	Val	Asp	Ala	Ala	Pro	Val	Pro	Pro	Ala	Pro				
			740						745						750				
Asp	Pro	Ala	Gly	Leu	Pro	Gly	Pro	Val	Val	Leu	Thr	Pro	Pro	Pro	Pro				
		755					760					765							
Pro	Pro	Val	His	Lys	Pro	Ser	Ile	Pro	Pro	Pro	Ser	Arg	Asn	Arg	Arg				
	770					775					780								
Leu	Leu	Tyr	Thr	Tyr	Pro	Asp	Gly	Ala	Lys	Val	Tyr	Ala	Gly	Ser	Leu				
	785				790						795				800				
Phe	Glu	Ser	Asp	Cys	Asp	Trp	Leu	Val	Asn	Ala	Ser	Asn	Pro	Gly	His				
				805					810					815					
Arg	Pro	Gly	Gly	Gly	Leu	Cys	His	Ala	Phe	Tyr	Gln	Arg	Phe	Pro	Glu				
			820					825					830						
Ala	Phe	Tyr	Pro	Thr	Glu	Phe	Ile	Met	Arg	Glu	Gly	Leu	Ala	Ala	Tyr				
		835					840					845							
Thr	Leu	Thr	Pro	Arg	Pro	Ile	Ile	His	Ala	Val	Ala	Pro	Asp	Tyr	Arg				
	850					855					860								
Val	Glu	Gln	Asn	Pro	Lys	Arg	Leu	Glu	Ala	Ala	Tyr	Arg	Glu	Thr	Cys				
	865				870						875				880				
Ser	Arg	Arg	Gly	Thr	Ala	Ala	Tyr	Pro	Leu	Leu	Gly	Ser	Gly	Ile	Tyr				
			885					890						895					
Gln	Val	Pro	Val	Ser	Leu	Ser	Phe	Asp	Ala	Trp	Glu	Arg	Asn	His	Arg				
		900						905					910						
Pro	Gly	Asp	Glu	Leu	Tyr	Leu	Thr	Glu	Pro	Ala	Ala	Asn	Trp	Phe	Glu				
		915					920					925							
Ala	Asn	Lys	Pro	Ala	Gln	Pro	Val	Leu	Thr	Ile	Thr	Glu	Asp	Thr	Ala				
	930					935						940							
Arg	Thr	Ala	Asn	Leu	Ala	Leu	Glu	Ile	Asp	Ala	Ala	Thr	Glu	Val	Gly				
	945				950						955				960				
Arg	Ala	Cys	Ala	Gly	Cys	Thr	Ile	Ser	Pro	Gly	Ile	Val	His	Tyr	Gln				
			965					970						975					
Phe	Thr	Ala	Gly	Val	Pro	Gly	Ser	Gly	Lys	Ser	Arg	Ser	Ile	Gln	Gln				
		980						985					990						
Gly	Asp	Val	Asp	Val	Val	Val	Val	Pro	Thr	Arg	Glu	Leu	Arg	Asn	Ser				
		995					1000					1005							
Trp	Arg	Arg	Arg	Gly	Phe	Ala	Ala	Phe	Thr	Pro	His	Thr	Ala	Ala	Arg				
	1010				1015						1020								
Val	Thr	Ile	Gly	Arg	Arg	Val	Val	Ile	Asp	Glu	Ala	Pro	Ser	Leu	Pro				

025	1030	1035	1040
Pro His Leu Leu Leu Leu His Met Gln Arg Ala Ser Ser Val His Leu	1045	1050	1055
Leu Gly Asp Pro Asn Gln Ile Pro Ala Ile Asp Phe Glu His Ala Gly	1060	1065	1070
Leu Val Pro Ala Ile Arg Pro Glu Leu Ala Pro Thr Ser Trp Trp Xaa	1075	1080	1085
Val Thr His Arg Cys Pro Ala Asp Val Cys Glu Leu Ile Arg Gly Ala	1090	1095	1100
Tyr Pro Lys Ile Gln Thr Thr Ser Arg Val Leu Arg Ser Leu Phe Trp	1105	1110	1115
Asn Glu Pro Ala Ile Gly Gln Lys Leu Val Xaa Thr Gln Ala Ala Lys	1125	1130	1135
Ala Ala Asn Pro Gly Ala Ile Thr Val His Glu Ala Gln Gly Ala Thr	1140	1145	1150
Phe Thr Glu Thr Thr Ile Ile Ala Thr Ala Asp Ala Arg Gly Leu Ile	1155	1160	1165
Gln Ser Ser Arg Ala His Ala Ile Val Ala Leu Thr Arg His Thr Glu	1170	1175	1180
Lys Cys Val Ile Leu Asp Ala Pro Gly Leu Leu Arg Glu Val Gly Ile	1185	1190	1195
Ser Asp Val Ile Val Asn Asn Phe Phe Leu Ala Gly Gly Glu Val Gly	1205	1210	1215
Xaa His Arg Pro Ser Val Ile Pro Arg Gly Asn Pro Asp Gln Asn Leu	1220	1225	1230
Gly Thr Leu Gln Ala Phe Pro Pro Ser Cys Gln Ile Ser Ala Tyr His	1235	1240	1245
Gln Leu Ala Glu Glu Leu Gly His Arg Pro Ala Pro Val Ala Ala Val	1250	1255	1260
Leu Pro Pro Cys Pro Glu Leu Glu Gln Gly Leu Leu Tyr Met Pro Gln	1265	1270	1275
Glu Leu Thr Val Ser Asp Ser Val Leu Val Phe Glu Leu Thr Asp Ile	1285	1290	1295
Val His Cys Arg Met Ala Ala Pro Ser Gln Arg Lys Ala Val Leu Ser	1300	1305	1310
Thr Leu Val Gly Arg Tyr Gly Arg Arg Thr Lys Leu Tyr Glu Ala Ala	1315	1320	1325
His Ser Asp Val Arg Glu Ser Leu Ala Arg Phe Ile Pro Thr Ile Gly			

1330	1335	1340
Pro Val Gln Ala Thr Thr Cys Glu Leu Tyr Glu Leu Val Glu Ala Met 345 1350 1355 1360		
Val Glu Lys Gly Gln Asp Gly Ser Ala Val Leu Glu Leu Asp Leu Cys 1365 1370 1375		
Asn Arg Asp Val Ser Arg Ile Thr Phe Phe Gln Lys Xaa Cys Asn Lys 1380 1385 1390		
Phe Thr Thr Gly Glu Thr Ile Ala His Gly Lys Val Gly Gln Gly Ile 1395 1400 1405		
Ser Ala Trp Ser Lys Thr Phe Cys Ala Leu Phe Gly Pro Trp Phe Arg 1410 1415 1420		
Ala Ile Glu Lys Glu Ile Leu Ala Leu Leu Pro Pro Asn Ile Phe Tyr 425 1430 1435 1440		
Gly Asp Ala Tyr Glu Glu Ser Val Phe Ala Ala Ala Val Ser Gly Ala 1445 1450 1455		
Gly Ser Cys Met Val Phe Glu Asn Asp Phe Ser Glu Phe Asp Ser Thr 1460 1465 1470		
Gln Asn Asn Phe Ser Leu Gly Leu Glu Cys Val Val Met Glu Glu Cys 1475 1480 1485		
Gly Met Pro Gln Trp Leu Ile Arg Leu Tyr His Leu Val Arg Ser Ala 1490 1495 1500		
Trp Ile Leu Gln Ala Pro Lys Glu Ser Leu Lys Gly Phe Trp Lys Lys 505 1510 1515 1520		
His Ser Gly Glu Pro Gly Thr Leu Leu Trp Asn Thr Val Trp Asn Met 1525 1530 1535		
Ala Ile Ile Ala His Cys Tyr Glu Phe Arg Asp Phe Arg Val Ala Ala 1540 1545 1550		
Phe Lys Gly Asp Asp Ser Val Val Leu Cys Ser Asp Tyr Arg Gln Ser 1555 1560 1565		
Arg Asn Ala Ala Ala Leu Ile Ala Gly Cys Gly Leu Lys Leu Lys Val 1570 1575 1580		
Asp Tyr Arg Pro Ile Gly Leu Tyr Ala Gly Val Val Val Ala Pro Gly 585 1590 1595 1600		
Leu Gly Thr Leu Pro Asp Val Val Arg Phe Ala Gly Arg Leu Ser Glu 1605 1610 1615		
Lys Asn Trp Gly Pro Gly Pro Glu Arg Ala Glu Gln Leu Arg Leu Ala 1620 1625 1630		
Val Cys Asp Phe Leu Arg Gly Leu Thr Asn Val Ala Gln Val Cys Val		

1635 1640 1645
 Asp Val Val Ser Arg Val Tyr Gly Val Ser Pro Gly Leu Val His Asn
 1650 1655 1660
 Leu Ile Gly Met Leu Gln Thr Ile Ala Asp Gly Lys Ala His Phe Thr
 665 1670 1675 1680
 Glu Thr Ile Lys Pro Val Leu Asp Leu Thr Asn Ser Ile Ile Gln Arg
 1685 1690 1695
 Val Glu

 <210> 92
 <211> 660
 <212> PRT
 <213> Hepatitis E virus

 <400> 92
 Met Arg Pro Arg Ala Val Leu Leu Leu Phe Leu Met Phe
 1 5 10
 Leu Pro Met Leu Pro Ala Pro Pro Ala Gly Gln Pro Ser Gly Arg Arg
 15 20 25
 Arg Gly Arg Arg Ser Gly Gly Ala Gly Gly Gly Phe Trp Ser Asp Arg
 30 35 40 45
 Val Asp Ser Gln Pro Phe Ala Leu Pro Tyr Ile His Pro Thr Asn Pro
 50 55 60
 Phe Ala Ala Asp Val Val Ser Gln Pro Gly Ala Gly Thr Arg Pro Arg
 65 70 75
 Gln Pro Pro Arg Pro Leu Gly Ser Ala Trp Arg Asp Gln Ser Lys Arg
 80 85 90
 Pro Ser Val Ala Pro Arg Arg Arg Ser Thr Pro Ala Gly Ala Ala Pro
 95 100 105
 Leu Thr Ala Ile Ser Pro Ala Pro Asp Thr Ala Pro Val Pro Asp Val
 10 115 120 125
 Asp Ser Arg Gly Ala Ile Leu Arg Arg Gln Tyr Asn Leu Ser Thr Ser
 130 135 140
 Pro Leu Thr Ser Ser Val Ala Ser Gly Thr Asn Leu Val Leu Tyr Ala
 145 150 155
 Ala Pro Leu Asn Pro Leu Leu Pro Leu Gln Asp Gly Thr Asn Thr His
 160 165 170
 Ile Met Ala Thr Glu Ala Ser Asn Tyr Ala Gln Tyr Arg Val Val Arg
 175 180 185
 Ala Thr Ile Arg Tyr Arg Pro Leu Val Pro Asn Ala Val Gly Gly Tyr

90	195	200	205
Ala Ile Ser Ile Ser Phe Trp Pro Gln Thr Thr Thr Thr Pro Thr Ser	210	215	220
Val Asp Met Asn Ser Ile Thr Ser Thr Asp Val Arg Ile Leu Val Gln	225	230	235
Pro Gly Ile Ala Ser Glu Leu Val Ile Pro Ser Glu Arg Leu His Tyr	240	245	250
Arg Asn Gln Gly Trp Arg Ser Val Glu Thr Thr Gly Val Ala Glu Glu	255	260	265
Glu Ala Thr Ser Gly Leu Val Met Leu Cys Ile His Gly Ser Pro Val	270	275	280
Asn Ser Tyr Thr Asn Thr Pro Tyr Thr Gly Ala Leu Gly Leu Leu Asp	290	295	300
Phe Ala Leu Glu Leu Glu Phe Arg Asn Leu Thr Pro Gly Asn Thr Asn	305	310	315
Thr Arg Val Ser Arg Tyr Thr Ser Thr Ala Arg His Arg Leu Arg Arg	320	325	330
Gly Ala Asp Gly Thr Ala Glu Leu Thr Thr Thr Ala Ala Thr Arg Phe	335	340	345
Met Lys Asp Leu His Phe Thr Gly Thr Asn Gly Val Gly Glu Val Gly	350	355	360
Arg Gly Ile Ala Leu Thr Leu Phe Asn Leu Ala Asp Thr Leu Leu Gly	370	375	380
Gly Leu Pro Thr Glu Leu Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr	385	390	395
Ser Arg Pro Val Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr	400	405	410
Thr Ser Val Glu Asn Ala Gln Gln Asp Lys Gly Ile Thr Ile Pro His	415	420	425
Asp Ile Asp Leu Gly Asp Ser Arg Val Val Ile Gln Asp Tyr Asp Asn	430	435	440
Gln His Glu Gln Asp Arg Pro Thr Pro Ser Pro Ala Pro Ser Arg Pro	450	455	460
Phe Ser Val Leu Arg Ala Asn Asp Val Leu Trp Leu Ser Leu Thr Ala	465	470	475
Ala Glu Tyr Xaa Gln Thr Thr Tyr Gly Ser Ser Thr Asn Pro Met Tyr	480	485	490
Val Ser Asp Thr Val Thr Leu Val Asn Val Ala Thr Gly Ala Gln Ala			

495 500 505
 Val Ala Arg Ser Leu Asp Trp Ser Lys Val Thr Leu Asp Gly Arg Pro
 10 515 520 525
 Leu Thr Thr Ile Gln Gln Tyr Ser Lys Lys Phe Tyr Val Leu Pro Leu
 530 535 540
 Xaa Gly Lys Leu Ser Phe Trp Glu Ala Gly Thr Thr Lys Ala Gly Tyr
 545 550 555
 Pro Tyr Asn Tyr Asn Thr Thr Ala Ser Asp Gln Ile Leu Ile Glu Asn
 560 565 570
 Ala Ala Gly His Arg Val Ala Ile Ser Thr Tyr Thr Thr Ser Leu Gly
 575 580 585
 Ala Gly Pro Thr Ser Xaa Ser Ala Val Gly Val Leu Ala Pro His Ser
 90 595 600 605
 Ala Leu Ala Val Leu Glu Asp Thr Val Asp Tyr Pro Ala Arg Ala His
 610 615 620
 Thr Phe Asp Asp Phe Cys Pro Glu Cys Arg Thr Leu Gly Leu Gln Gly
 625 630 635
 Cys Ala Phe Gln Ser Thr Ile Ala Glu Leu Gln Arg Leu Lys Met Lys
 640 645 650
 Val Gly Lys Thr Arg Glu Ser
 655 660

 <210> 93
 <211> 122
 <212> PRT
 <213> Hepatitis E virus

 <220>
 <223> ORF3 HEV US-1

 <400> 93
 Met Asn Asn Met Ser Phe Ala Ser Pro Met Gly Ser Pro Cys Ala Leu
 1 5 10 15
 Gly Leu Phe Cys Cys Cys Ser Ser Cys Phe Cys Leu Cys Cys Pro Arg
 20 25 30
 His Arg Pro Val Ser Arg Leu Ala Val Ala Val Gly Gly Ala Ala Ala
 35 40 45
 Val Pro Ala Val Val Ser Gly Val Thr Gly Leu Ile Leu Ser Pro Ser
 50 55 60
 Pro Ser Pro Ile Phe Ile Gln Pro Thr Pro Ser Pro Pro Met Ser Phe
 65 70 75 80

His Asn Pro Gly Leu Glu Leu Ala Leu Asp Ser Arg Pro Ala Pro Ser
 85 90 95

Val Pro Leu Gly Val Thr Ser Pro Ser Ala Pro Pro Leu Pro Pro Val
 100 105 110

Val Asp Leu Pro Gln Leu Gly Leu Arg Arg
 115 120

<210> 94
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer
 US5P3S/20

<400> 94
 tggcattact actgccattg 20

<210> 95
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer
 US5P45S/20

<400> 95
 caattctgcc ttggcgaatg 20

<210> 96
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer
 US5P296A

<400> 96
 aggaaacacc gatgcagaac 20

<210> 97
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer
 US5P243A/20

<400> 97
tccaacctcc aagcaacgac 20

<210> 98
<211> 199
<212> DNA
<213> Hepatitis E virus

<220>
<223> Clone 199con

<400> 98
caattctgcc ttggcgaatg ctgtggtggt tgggcggtt ctttctcgtg tgcaaactga 60
gattcttatt aatttgatgc aaccccgga gttggtcttc cgccctgagg tgctttggaa 120
tcctctatc cagcgggtta tacataatga attagagcag tactgccggg cccgggctgg 180
tcgttgcttg gaggttga 199

<210> 99
<211> 25
<212> DNA
<213> Hepatitis E virus

<220>
<223> JE orf1-s

<400> 99
gttctgcac ggtgtttcct tagac 25

<210> 100
<211> 26
<212> DNA
<213> Hepatitis E virus

<220>
<223> JE orf1-a

<400> 100
gaatcaggag atacgaggtt gtgtgg 26

<210> 101
<211> 331
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-320

<400> 101
gttctgcac ggtgtttcct tagaccgggc ggccgagatg ttcagcgctg gtattctgcc 60

cctacccgtg gtccctgcggc caattgccgc cgctccgagt tgcgtggtct ccccccctgtc 120
 gaccgcacct attgttttga tggattttcc cgttgtgctt ttgctgcaga gaccgggtgtg 180
 gccctttact ctttgcata cctttggcca gctgatgttg cagaggctat ggcccgcctat 240
 gggatgacac gcttatacgc cgcactgcac cttccccccg aggtgctgct accaccgggc 300
 acctaccaca caacctcgta tctcctgatt c 331

<210> 102

<211> 1186

<212> DNA

<213> Hepatitis E virus

<220>

<223> us2-1168

<400> 102

ctccactgtgt ccgatagtgt gttggttttt gagcttacgg atatagtcca ctgccgtatg 60
 gccgccccaa gccagcgaaa ggctgtttct tcaacgcttg tggggaggta cggccgtagg 120
 actaaattat atgaggcggc gcattcagat gtccgtgagt ccctagcgag gtttatcccc 180
 accatcgggc ctgttcgggc taccacatgt gagctgtacg agctggttga agccatggta 240
 gagaagggtc aggacggatc tgccgtccta gagctcgacc tttgcaatcg tgacgtctcg 300
 cgcatacat ttttccaaaa ggattgcaat aagtttataa ctggtgagac tatcgcccat 360
 ggcaagggtg gccagggcat atcggcctgg agcaagacct tctgtgctct gtttggcccc 420
 tggttccgcy ccattgaaaa ggaaatattg gccctactcc cgccataat cttttatggc 480
 gacgcctatg aggagtcagt gtttgctgcc gctgtgtccg gggcagggtc atgtatggta 540
 tttgaaaatg acttctcaga gtttgacagt acccagaata atttctctct cggccttgag 600
 tgtgtggtta tggaggagt cggcatgccc caatgggtta ttaggttgta ccatctggtc 660
 cggtcagcct ggatttttga gccgccgaag gactctctta aggggttttg gaagaagcac 720
 tctggtgagc ctggtaccct tctctggaac actgtctgga acatggcgat tatagcacat 780
 tgctaygagt tccgtgactt tctgtgtgcc gccttcaagg gtgatgattc agtggctctc 840
 tgtagtgact accgacagrg ccgtaacgcy gctgccttaa ttgcaggctg tgggctcaaa 900
 ttgaagggtg attaccgccc tatcgggcta tatgctggag tggtggtggc ccccggtttg 960
 gggacactgc ccgatgtggt gcgttttgcc ggtcgggttat ctgagaagaa ttggggccct 1020
 ggcccgagcy gtgctgagca gctgcgtctt gctgtttgtg atttccttcg aggggtgacg 1080

aatgttgccg aggtctgtgt tgatgttggtg tcccggtgt atggagttag ccccggtgtg 1140
 gtacataacc ttattggcat gctgcagacc atcgccgatg gcaagg 1186

<210> 103
 <211> 23
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> JE hevdf2/3 s1

<400> 103
 gttccgcttg gcgtgaccag tcc 23

<210> 104
 <211> 23
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> JE hevdf2/3 a1

<400> 104
 gagtcaacat caggtacagg agc 23

<210> 105
 <211> 130
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-135

<400> 105
 gttccgcttg gcgtgaccag tccagcgcc cctccgctgc ccccgctcgt cgatctgccc 60
 cagctggggc tgcgccgtg actgccgtgt caccggtcc tgacacagct cctgtacctg 120
 atgttgactc 130

<210> 106
 <211> 26
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> JE hevdf1-s1

<400> 106
 gatgtcattt tgtgttgctg ctcacc 26

<210> 107
 <211> 23
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> hev216 a1

<400> 107
 cgtcctacag cggcatggta ctg

23

<210> 108
 <211> 564
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-563

<400> 108
 tcacccatgc cttatgttcc ttaccctcgt tcaacggagg tgtatgtccg gtctatatatt 60
 ggccctggcg gctccccatc cttgtttcca tcagcctgct ctactaaatc tacctttcat 120
 gctgtcccg ttcacatctg ggatcrgctc atgctctttg gtgccaccct gracgatcag 180
 gcgttctgct gttcaccggt tatgacttac ctccgtggta ttagttataa ggtcactgtc 240
 ggtgcgcttg tcgctaata ggggtggaac gcctctgagg atgctcttac tgcagtgatc 300
 actgcggcct atctgacat ctgccatcag cgttaccttc gcacccagge gatttccaag 360
 ggcatgcgcc ggttgagggt tgagcatgct cagaaattta tcacaagact ctacagctgg 420
 ctatttgaga agtctggccg tgactacatc cccggccgcc agcttcaatt ttatgcacaa 480
 tgccgacggt ggctttctgc aggttccac ctaraccca gggtgcttgt ctttgatgaa 540
 tcagtaccat gccgctgtag gacg 564

<210> 109
 <211> 24
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> USorf2.1'

<400> 109
 gtggagctag tacaccgacc gcag

24

<210> 110
 <211> 678
 <212> DNA

<213> Hepatitis E virus

<220>

<223> us2-667

<400> 110

```
cgcttcttgg tggtttaccg acagaattga tttcgtcggc tgggggccaa ctgttttact 60
cccgcccggg tgtctcagcc aatggcgagc caacagtaaa gttatataca tctgttgaga 120
atgcgcgagc agacaagggc atcaccattc cacatgatat agacctgggt gactcccggt 180
tggttatcca ggattatgat aaccagcayg agcaagaccg acctactccg tcacctgccc 240
cctctcgccc cttctcagtt cttcgtgcc aatgatgttt gtggctttcc ctcactgccg 300
ctgagtatga ccagactacg tatgggtcgt ccaccaaccc tatgtatgtc tctgacacag 360
ttacgcttgt taatgtggct actgggtgct aggctgttgc ccgctccctt gattgggtcta 420
aagttactct ggacggccgc ccccttacta ccattcagca gtattctaag acattttatg 480
ttctcccgtc ccgcggaag ctgtcctttt gggaggctgg cactactaag gccggctacc 540
cttacaatta taatactacc gctagtgacc aaattttgat tgagaatgag gccggccacc 600
gtgtcgctat ttccacctat accactagct taggtgccgg tcctacctcg atctctgcgg 660
tcggtgtact agctccac
```

678

<210> 111

<211> 23

<212> DNA

<213> Hepatitis E virus

<220>

<223> hev3301s

<400> 111

gtatgcgagc tcatccgtgg tgc

23

<210> 112

<211> 25

<212> DNA

<213> Hepatitis E virus

<220>

<223> JE hev167-a1

<400> 112

ccaacacact atcggacaca gtgag

25

<210> 113

<211> 580

<212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-579

<400> 113
 gtatgcgagc tcatccgtgg tgcctacccc aaaattcaga ccacgagccg tgtgctacgg 60
 tccctgtttt ggaacgaacc ggccatcggc caaaagttgg tttttacgca ggctgctaag 120
 gctgccaaacc ctggtgcgat tacggttcac gaagctcagg gtgctacttt cacggagacc 180
 acaattatag ccacggccga cgctaggggc ctcattcagt catcccgggc ccatgctata 240
 gtcgcactca cccgccatac tgagaagtgt gttattttgg atgccccggg cttgttgccg 300
 gaggtcggca tttcggatgt tattgtcaat aactttttcc ttgccggtgg agaggtcggc 360
 catcaccgcc cttctgtgat acctcgcggc aatcctgac agaacctcgg gactctacag 420
 gcctttccgc cgtcatgtca gatcagtgtc taccatcagt tggctgagga actagggtcat 480
 cgcccgcccc ctgtcgccgc cgtcttgccc ccttgcccctg agcttgagca gggcctgctc 540
 tatatgccac aagaactcac tgtgtccgat agtgtgttgg 580

<210> 114
 <211> 26
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> HEV459 s1

<400> 114
 cagaaattta tcacaagact ctacag 26

<210> 115
 <211> 26
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> HEV459 s3

<400> 115
 ctctacagtt ggctatttga gaagtc 26

<210> 116
 <211> 25
 <212> DNA
 <213> Hepatitis E virus

<220>

<223> JE1955a

<400> 116

ctataaagag ctgagcagaa ggcgg

25

<210> 117

<211> 734

<212> DNA

<213> Hepatitis E virus

<220>

<223> us2-733

<400> 117

ctctacagtt ggctatttga gaagtctggc cgtgactaca tccccggcgg ccagcttcaa 60
 ttttatgcac aatgccgacg gtgggtttct gcaggettcc acctaraccc caggrtgctt 120
 gtctttgatg aatcagtgcc atgccgttgc aggacgtttt tgaagaaggc cgcgggtaaa 180
 ttctgtctgt ttatgcgggtg gctggggcag gagtgtacct gcttcttggg gccagccgag 240
 ggttttagttg gtgatcaagg tcatgacaac gaggcctatg aaggttctga ggtcgaccca 300
 gctgagcctg cacatcttga tgtctcgggg acttatgccg tccatgggca ccagcttgag 360
 gccctctata gggcacttaa tgtcccacat gatattgccg ctcgagcctc ccgactaacg 420
 gctactgttg agctcgttgc tagtccggac cgcttagagt gccgcactgt acttggtaat 480
 aagaccttcc ggacgacggt ggttgatggc gcccatcttg aagcgaatgg ccctgaggag 540
 tatgttctgt catttgacgc ctctcgccag tctatggggg ccgggtcgca cagcctcact 600
 tatgagctca cccctgccgg tctgcaggta aagatttcat ctaatggtct ggattgcact 660
 gccacattcc ccccyggtgg cgcacctagc gccgcgccgg gggaggtggc cgccttctgc 720
 tcagctcttt atag 734

<210> 118

<211> 22

<212> DNA

<213> Hepatitis E virus

<220>

<223> JE 2950mex s

<400> 118

gtgtccccgg ctctggcaag tc

22

<210> 119

<211> 22

<212> DNA
 <213> Hepatitis E virus

<220>
 <223> JE us2-579-a2

<400> 119
 cagggttggc agccttagca gc 22

<210> 120
 <211> 483
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-482

<400> 120
 gtgtccccgg ctctggcaag tcaaggtcca tacaacaggg agatgtcgat gtggtggttg 60
 tgcccacccg ggagctccgt aacagctggc gtcgcggggg ttttgcggcc ttcacacctc 120
 acacagcggc ccgtgttact atcggcggcc gcgttgatgat tgatgaggct ccattctctc 180
 caccgcacct gctgctgtta cacatgcagc gggcctcttc ggtccatctc cttggtgatc 240
 caaaccagat tcctgctatt gatcttgagc atgcgggctt ggtccccgag atccgccccg 300
 agcttgcgcc aacgagctgg tggcacgtta cacaccgttg cccggccgat gtgtgcgagc 360
 tcatacgtgg ggcctacccc aaaattcaga ccaagagccg tgtgctacgg tcctgtttt 420
 ggaacgaacc ggccatcggc caaaagttgg tttttacgca ggctgctaag gctgccaacc 480
 ctg 483

<210> 121
 <211> 24
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> JE 2600s

<400> 121
 taacccaaag aggcttgagg ctgc 24

<210> 122
 <211> 22
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-482-a1

<400> 122
ccgctgtgtg aggtgtgaag gc

22

<210> 123
<211> 23
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-482-a2

<400> 123
gacgccagct gttacggagc tcc

23

<210> 124
<211> 431
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-430

<400> 124
taaccctaaag aggcttgagg ctgcgtaccg ggaaacttgc tcccgctgtg gcaccgctgc 60
ctaccgcttt ttgggctcgg gtatatacca ggtccctgtt agcctcagtt ttgatgcctg 120
ggaacgcaat caccgccccg gcgatgagct ttacttgaca gagcccgccg cagcctgggt 180
tgaggctaata aagccggcgc agccggcgct tactataact gaggacacgg cccgtacggc 240
caacctggca ttagagattg atgccgccac agaggttggc cgtgcttgtg ccggctgcac 300
catcagcccc gggatttgtc actatcagtt taccgccccg gtccccgggt caggcaagtc 360
aaggccata caacaggag atgtcgatgt ggtggttgtg cccaccggg agctccgtaa 420
cagctggcgt c 431

<210> 125
<211> 22
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-orf2/3 s1

<400> 125
cgtcgtgat ctgccccagc tg

22

<210> 126
<211> 25

<212> DNA
 <213> Hepatitis E virus

<220>
 <223> HEVConsORF2-a1

<400> 126
 cttgttctrtg ytggttrtca taatc 25

<210> 127
 <211> 21
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-orf2/3 s2

<400> 127
 cgctgactgc cgtgtcacccg g 21

<210> 128
 <211> 25
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> HEVConsORF2-a2

<400> 128
 gttcrtgytg gttrtcataa tcctg 25

<210> 129
 <211> 1020
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-1019

<400> 129
 cgctgactgc cgtgtcacccg gctcctgaca cagcccctgt acctgatgtt gactcacgtg 60
 gtgctattct gcgccggcag tacaatttgt ccacgtcccc gctcacgtca tetgtcgctt 120
 cgggtactaa tttggtcctc tatgctgcc cgctgaatcc cctcttgcc ctccaggatg 180
 gtaccaaacac tcatattatg gctactgagg catccaatta tgcccagtat cgggttggtc 240
 gagctacaat ccgttatcgc ccgctgggtgc cgaatgccgt tgggtggetat gccatttcca 300
 tttctttctg gccccaaact acaactaccc ctactttctgt cgatatgaat tctattactt 360
 ccacygatgt taggattttg gttcagcccc gtattgcctc cgagctagtc atccccagtg 420

agcgcccttca ttaccgtaat caaggctggc gctctgttga gaccacgggt gtggctgagg 480
 aggaggctac ttccggtctg gtaatgcttt gcattcatgg ctctcctggt aattcctaca 540
 ctaatacacc ttacactggg gcgctggggc ttcttgattt tgcactagag cttgaattta 600
 ggaatttgac acccggaac accaacaccc gtgtttcccg gtataccagc acagcccgcc 660
 accggctgcg ccgtgggtgct gatgggactg ctgagcttac taccacagca gccacacgtt 720
 tcatgaagga cctgcacttc gctggcacga atggcggttg tgaggtgggt cgtgggtatcg 780
 ccctgacact gttcaatctc gctgatacgc ttctcggcgg tttaccgaca gaattgattt 840
 cgtcggctgg gggccaactg ttttactccc gcccggttgt ctacagccaat ggcgagccaa 900
 cagtaaagtt atatacatct gttgagaatg cgcagcaaga caagggcatc accattccac 960
 atgatataga cctgggtgac tcccggtggt ttatccagga ttatgataac cagcaygaac 1020

<210> 130
 <211> 24
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2 330s1

<400> 130
 cagctgatgt tgcagaggct atgg

24

<210> 131
 <211> 24
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2 563a1

<400> 131
 gcaggctgat ggaaacaagg atgg

24

<210> 132
 <211> 407
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-406

<400> 132
 cagctgatgt tgcagaggct atggcccgcc atgggatgac acgcttatac gccgcactgc 60
 accttcccc cgaggtgctg ctaccacccg gcacctacca cacaacctcg tacctcttga 120

ttcacgatgg caaccgcgct gttgtaactt acgagggcga tactagtgcg ggctataatc 180
 atgatgtctc catacttcgt gcatggatcc gtactactaa aatagttggg gaccatccat 240
 tggatcataga gcgagtgcgg gccattgggt gtcattttgt gctgctgctc accgcagccc 300
 ctgaaccgtc acctatgcct tatgttcctt accctcggtc aacggagggt tatgtccggg 360
 ctatatttgg ccttggcggc tccccatcct tgtttccatc agcctgc 407

<210> 133
 <211> 22
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-579 s1

<400> 133
 cagaccacga gccgtgtgct ac 22

<210> 134
 <211> 23
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-1168 a1

<400> 134
 ccacaagcgt tgagagaaca gcc 23

<210> 135
 <211> 22
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-579 s2

<400> 135
 gctgctaagg ctgccaaccc tg 22

<210> 136
 <211> 547
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-579wb

<400> 136

gctgctaagg ctgccaaccc tgggtgcgatt acggttcacg aagctcaggg tgctactttc 60
acggagacca caattatagc cacggccgac gctagggggc tcattcagtc atcccggggc 120
catgctatag tcgcactcac ccgccatact gagaagtgtg ttattttgga tgcccccggc 180
ttgttgcgcg aggtcggcat ttcggatggt attgtcaata actttttcct tgccggtgga 240
gaggtcggcc atcacccgcc ttctgtgata cctcgcggca atcctgatca gaacctcggg 300
actctacagg cttttccgcc gtcatgtcag atcagtgtt accatcagtt ggctgaggaa 360
ctaggtcatc gcccgcccc tgtcgccgcc gtcttgcccc cttgccctga gcttgagcag 420
ggcctgctct atatgccaca agaacttact gtgtccgata gcgtgctggt ttttgagctt 480
acggatatag tccactgccg tatggccgcc ccaagccagc gaaaggctgt tctctcaacg 540
cttgtgg 547

<210> 137
<211> 24
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-733s1

<400> 137
cacagcctca cttatgagct cacc 24

<210> 138
<211> 23
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-430a1

<400> 138
cggtgattgc gttcccaggc atc 23

<210> 139
<211> 26
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-733s2

<400> 139
ctgcaggtaa agatttcacg taatgg 26

<210> 140
 <211> 24
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-430a2

<400> 140
 ccaggcatca aaactgaggc taac

24

<210> 141
 <211> 903
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-851

<400> 141
 ctgcaggtaa agatttcac taatgggtctg gattgcactg ccacattccc cccygggtggc 60
 gcccttagcg ccgcgccggg ggaggtggcs gccttctgca gtgctcttta tagatacaat 120
 aggttcaccc agcggcattc gctgacaggc ggactatggc tacatcctga ggggctgctg 180
 ggtatcttcc cccattctc ccctgggcat atttgggagt ctgctaacc cttttgcggg 240
 gaggggactt tgtatacccg aacctggtca acctctggtt tttctagtga tttctcccc 300
 cctgaggcgg ccgtcctcgc ttcggtgcc gccccggggg tgcctaccc tactccacct 360
 gttagtata tctgggtgtt accaccgcc tcagaggaat ctcatgttg tgccgcatct 420
 gtacctctg ttctgagcc tgcgtgattg accagcccta ttgtgcttac ccccccccc 480
 cccctctc ccgtgcgtaa gccggcaaca tcccgcctc ccgcactcg ccgtctcctt 540
 tacacctacc ccgacggcgc caaggtgtat gcggggcat tgtktagtc agactgtgat 600
 tggttagtca atgcctcaaa ccctggccat cgccccgggg gtggcctctg ccatgctttt 660
 tatcaacgtt tcccagaagc gttctactcg actgaattca tcatgcgcga gggccttgca 720
 gcatacactt taaccccgcg ccctattatc catgcagtgg ctcccgacta tagggttgag 780
 caaaaccgga agaggcttga ggcagcgtac cgggaaactt gctcccgtcg tggcaccgct 840
 gcctaccgca ttttgggctc gggatatata caggtccctg ttagcctcag ttttgatgcc 900
 tgg

903

<210> 142
 <211> 24
 <212> DNA

<213> Hepatitis E virus

<220>

<223> us2-1168s1

<400> 142

gcaggtctgt gttgatgttg tgtc

24

<210> 143

<211> 21

<212> DNA

<213> Hepatitis E virus

<220>

<223> us2-dforf2/3 a2

<400> 143

ccggtgacac ggcagtcagc g

21

<210> 144

<211> 25

<212> DNA

<213> Hepatitis E virus

<220>

<223> us2-1168s2

<400> 144

gatgttgtgt cccgtgtcta tggag

25

<210> 145

<211> 22

<212> DNA

<213> Hepatitis E virus

<220>

<223> us2 dforf2/3 a3

<400> 145

cagctggggc agatcgacga cg

22

<210> 146

<211> 503

<212> DNA

<213> Hepatitis E virus

<220>

<223> us2-502

<400> 146

gatgttgtgt cccgtgtcta tggagttagc cccgggctgg tacataacct tattggcatg 60

ctgcagacca ttgctgatgg caaggccac ttacagara atattaaacc tgtgcttgac 120

cttacaaatt ccatcataca acgggtggaa tgaataacat gtcttttgca tcgcccattg 180
 gatcaccatg cgccctaggg ctgttctgtt gttgctcttc gtgcttttgc ctatgctgcc 240
 cgcgccaccg gccggccagc cgtctggcgg ccgtcgtggg cggcgcagcg gcggtgccgg 300
 cggtggtttc tggggtgaca gggttgattc tcagcccttc gccctcccct atattcatcc 360
 aaccaacccc ttcgccgcgg atgtcgtttc acaaccggg gctggaactc gccctcgaca 420
 gccgccccgc ccccttgggt ccgcttggcg tgaccagtc cagcgccct ccgctgcccc 480
 ccgtcgtcga tctgccccag ctg 503

<210> 147

<211> 24

<212> DNA

<213> Hepatitis E virus

<220>

<223> HEVConsORF1-s1

<400> 147

ctggcatyac tactgcyatt gagg 24

<210> 148

<211> 23

<212> DNA

<213> Hepatitis E virus

<220>

<223> HEVConsORF1-a1

<400> 148

ccatcrarrc agtaagtgcg gtc 23

<210> 149

<211> 418

<212> DNA

<213> Hepatitis E virus

<220>

<223> us2-orf1

<400> 149

ctggcattac tactgctatt gaggagctg ctctggctgc ggctaattcc gccttggcga 60
 atgctgtggt ggttcggcgg tttctttctc gtgtgcaaac tgagattctt attaatttga 120
 tgcaaccccg gcagttggtc ttccgccctg aggtgctttg gaatcactct atccagcggg 180
 ttatacataa tgaattagag cagtactgcc gggcccgggc tggctgttgt ttggaggttg 240

gagccacccc gaggtccatt aatgacaacc ctaatgtctt gcataggtgt tttcttagac 300
 cggtcggccg agatgttcag cgctgggtatt ctgcccctac ccgtgggtcct gcggccaatt 360
 gccgcccgtc cgcgttgctt ggtctccccc ctgtcgaccg cacttactgt tttgatgg 418

<210> 150
 <211> 24
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> HEVConsORF2-s1

<400> 150
 gacagaattr atttcgtcgg ctgg 24

<210> 151
 <211> 197
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-orf2

<400> 151
 gacagaattg atttcgtcgg ctggggggcca actgttttac tcccgcccgg ttgtctcagc 60
 caatggcgag ccaacagtaa agttatatac atctgttgag aatgcgcagc aagacaaggg 120
 catcaccatt ccacatgata tagacctggg tgactcccggt gtgggttatcc aggattatga 180
 taaccagcay gagcaag 197

<210> 152
 <211> 22
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> HEVConsORF2-s2

<400> 152
 gtygtctcrg ccaatggcga gc 22

<210> 153
 <211> 901
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-3p

<400> 153
gttggtctcag ccaatggcga gccaacagta aagttatata catctgttga gaatgcgcag 60
caagacaagg gcatcaccat tccacatgat atagacctgg gtgactcccc tgtgggttatc 120
caggattatg ataaccagca ygagcaagac cgacctactc cgtcacctgc cccctctcgc 180
cccttctcag ttcttcgtgc caatgatgtt ttgtggcttt ccctcactgc cgctgagtat 240
gaccagacta cgtatgggtc gtccaccaac cctatgtatg tctctgacac agttacgctt 300
gttaatgtgg ctactgggtc tcaggctgtt gcccgcctcc ttgattggtc taaagttact 360
ctggacggcc gcccccttac taccattcag cagtattcta agacatttta tgttctcccg 420
ctccgcggga agctgtcctt ttgggaggct ggcacgacta aggcgggcta cccttacaat 480
tataatacta ccgctagtga ccaaattttg attgagaatg cggccggcca ccgtgtcgct 540
atttccacct ataccactag cttagggtgcc ggtcctacct cgatctctgc ggtcggcgta 600
ctggctccac actctgccct tgccgttctt gaggatacta ttgattaccc cgcccggtgcc 660
catacttttg atgatttttg cccggagtgc cgtaccctag gtttgcaggg ttgtgcattc 720
cagtctacta ttgctgagct ccagcgttta aaaatgaagg taggtaaaac ccgggagtct 780
taattaattc cttctgtgcc cccttcgtag tttctttcgc ttttatttct tatttctgct 840
ttccgcgctc cctggaaaaa aaaaaaaaaa aaaaaaaaaa agtactagtc gacgcgtggc 900
c 901

<210> 154
<211> 27
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-gap s1

<400> 154
tatagataac aataggttca cccagcg 27

<210> 155
<211> 25
<212> DNA
<213> Hepatitis E virus

<220>
<223> us2-gap a1

<400> 155
attcagtcga gtagaacgct tctgg 25

<210> 156
 <211> 23
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-gap s2

<400> 156
 cggactatgg ctacatcctg agg 23

<210> 157
 <211> 26
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> us2-gap a2

<400> 157
 ttgactaacc aatcacagtc tgactc 26

<210> 158
 <211> 462
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> 13906-gap

<400> 158
 cggactatgg ctacatcctg aggggctgct gggtatcttc ccccatctct cccctgggca 60
 tattttgggag tctgctaacc cctttttgcgg tgaggggact ttgtataccc gaacctgggc 120
 aacctctggg ttttctagtg atttctcccc cctgaggcg gccgctcctg cttcgggctgc 180
 cgccccgggg ttgcctacc ctactccacc tgtagtgat atctgggtgt taccaccgcc 240
 ctgagaggaa tctcatgttg atgcggcatc tgtacctct gttcctgagc ctgctggatt 300
 gaccagcct attgtgctta ccccccccc cccctcct cccgtgcgta agccggcaac 360
 atccccgcct cccgcactc gccgtctcct ttacacctac cccgacggcg ccaaggtgta 420
 tgcgggggtca ttgtttgagt cagactgtga ttggttagtc aa 462

<210> 159
 <211> 21
 <212> DNA
 <213> Hepatitis E virus

<220>

<223> us-575a

<400> 159

gccgggtggt agcagcacct c

21

<210> 160

<211> 24

<212> DNA

<213> Hepatitis E virus

<220>

<223> us-426s

<400> 160

cgttgtgctt ttgctgcaga gacc

24

<210> 161

<211> 22

<212> DNA

<213> Hepatitis E virus

<220>

<223> us-84a

<400> 161

gaaacggccg aaccaccaca gc

22

<210> 162

<211> 24

<212> DNA

<213> Hepatitis E virus

<220>

<223> us-484s

<400> 162

cagctgatgt tgcagaggct atgg

24

<210> 163

<211> 22

<212> DNA

<213> Hepatitis E virus

<220>

<223> us-78a

<400> 163

gccgaaccac cacagcattc gc

22

<210> 164

<211> 7277

<212> DNA

<213> Hepatitis E virus

<220>

<223> us2full

<400> 164

tgcacagggg gcagaccacg tatgtggtcg atgccatgga ggcccatcag ttcattaagg 60
ctcctggcat tactactgct attgagcagg ctgctctggc tgcggctaatt tccgccttgg 120
cgaatgctgt ggtggttcgg ccgtttcttt ctgctgtgca aactgagatt cttattaatt 180
tgatgcaacc ccggcagttg gtcttccgcc ctgaggtgct ttggaatcat cctatccagc 240
gggttataca taatgaatta gagcagtact gccgggcccg ggctggtcgt tgtttggagg 300
ttggagccca cccgaggtcc attaatgaca accctaattgt cttgcatagg tgttttctta 360
gaccggtcgg ccgagatggt cagcgtggtt attctgcccc taccgctggg cctgcggcca 420
attgccgccc ctccgcgttg cgtggtctcc cccctgtcga ccgcacctat tgttttgatg 480
gattttcccg ttgtgctttt gctgcagaga ccggtgtggc cctttactct ttgcatgacc 540
tttggccagc tgatgttgca gaggetatgg ccgccatgg gatgacacgc ttatacgccg 600
cactgcacct tcccccgag gtgctgtac caccggcac ctaccacaca acctcgtacc 660
tcttgattca cgatggcaac cgcgtgttg taacttaoga gggcgatact agtgcgggct 720
ataatcatga tgtctccata ctctgtgcat ggatccgtac tactaaaata gttggtgacc 780
atccattggt catagagcga gtgcgggcca ttgggtgtca ttttgctgctg ctgctcaccg 840
cagcccctga accgtcacct atgccttatg ttccctacct tcgttcaacg gaggtgatg 900
tccggtctat atttggccct ggcggctccc catccttggt tccatcagcc tgctctacta 960
aatctacctt tcatgctgtc ccggttcaca tctgggater gctcatgctc tttggtgcca 1020
ccctgracga tcaggcgttc tgctgttcac ggcttatgac ttacctcgtt ggtattagtt 1080
ataaggtcac tgtcggtgcg cttgtcgcta atgaggggtg gaacgcctct gaggatgctc 1140
ttactgcagt gatcactgcg gcctatctga ccatctgcca tcagcgttac cttcgcaccc 1200
aggcgatttc caagggcatg cgccggttg aggttgagca tgcctcagaaa tttatcacia 1260
gactctacag ctggctatct gagaagtctg gccgtgacta catccccggc cgccagcttc 1320
aattttatgc acaatgccga cgggtggctt ctgcaggctt ccacctarac cccaggrtgc 1380
ttgtctttga tgaatcagtg ccatgccgtt gcaggacgtt tttgaagaag gtcgcgggta 1440
aattctgctg ttttatgcgg tggctggggc aggagtgtac ctgcttcttg gagccagccg 1500
agggttttagt tggtgatcaa ggtcatgaca acgaggccta tgaaggttct gaggtcgacc 1560

cagctgagcc tgcacatctt gatgtctcgg ggacttatgc cgtccatggg caccagcttg 1620
aggccctcta tagggcactt aatgtcccac atgatattgc cgctcgagcc tcccactaa 1680
cggctactgt tgagctcgtt gctagtccgg accgcttaga gtgccgcact gtacttggtg 1740
ataagacctt ccggacgacg gtggttgatg gcgcccatct tgaagcgaat ggccctgagg 1800
agtatgttct gtcatttgac gcctctcgcc agtctatggg ggccgggtcg cacagcctca 1860
cttatgagct caccctcgcc ggtctgcagg taaagatttc atctaattgg ctggattgca 1920
ctgccacatt ccccccyygt ggccccccta gcgcgcgcgc gggggagggtg gcsgecttct 1980
gcagtgtctt ttatagatac aataggttca cccagcggca ttcgctgaca ggccggactat 2040
ggctacatcc tgaggggctg ctgggtatct tccccccatt ctccctggg catatttggg 2100
agtctgctaa ccccttttgc ggtgagggga ctttgtatac ccgaacctgg tcaacctctg 2160
gtttttctag tgatttctcc cccctgagg cggccgctcc tgettoggct gccgccccgg 2220
ggttgcccta ccctactcca cctgttagtg atatctgggt gttaccacgg ccctcagagg 2280
aatctcatgt tgatgcggca tctgtaccct ctgttctga gcctgctgga ttgaccagcc 2340
ctattgtgct tcccccccc cccccctc ctcccgctgg taagccggca acatccccgc 2400
ctccccgcac tcgccgtctc ctttacacct accccgacgg ccgaagggtg tatgcggggg 2460
cattgtktga gtcagactgt gattggttag tcaatgcctc aaacctggc catcgccccg 2520
gggggtggcct ctgccatgct ttttatcaac gtttccaga agcgttctac tcgactgaat 2580
tcatcatgcg cgagggcctt gcagcataca ctttaacccc gcgccctatt atccatgcag 2640
tggctcccga ctatagggtt gagcaaaacc cgaagaggct tgaggcagcg taccgggaaa 2700
cttgctcccc tcgtggcacc gctgcctacc cgcttttggg ctgggtata taccaggctc 2760
ctgttagcct cagttttgat gcctgggaac gcaatcacgg ccccgcgat gagctttact 2820
tgacagagcc cgccgcagcc tggtttgagg ctaataagcc ggccgcagcc gcgcttacta 2880
taactgagga cagggccgtt acggccaacc tggcattaga gattgatgcc gccacagagg 2940
ttggccgtgc ttgtgccggc tgcaccatca gcccgggat tgtgcactat cagtttaccg 3000
ccgggggtccc gggctcaggc aagtcaaggc ccatacaaca gggagatgtc gatgtggtgg 3060
ttgtgcccac ccgggagctc cgtaacagct ggctgcggc gggttttgcg gccttcacac 3120
ctcacacagc ggcccggtgt actatcgcc gccgcgttgt gattgatgag gctccatctc 3180
tcccacgca cctgctgctg ttacacatgc agcgggcctc ctcggtccat ctcttggtg 3240

atccaaacca gattcctgct attgattttg agcatgccgg cctgggtcccc gcgatccgcc 3300
 ccgagcttgc gccaacgagc tgggtggcacg ttacacaccg ttgcccggcc gatgtgtgcg 3360
 agctcatacy tggggcctac cccaaaatto agaccacgag ccgtgtgcta cgggccctgt 3420
 tttggaacga accggccatc ggccaaaagt tggtttttac gcaggctgct aaggctgcca 3480
 accctgggtgc gattacggtt cacgaagctc aggggtgctac tttcacggag accacaatta 3540
 tagccacggc cgacgctagg ggccctcatc agtcatcccg ggcccatgct atagtcgcac 3600
 tcacccgcca tactgagaag tgtgttattt tggatgcccc cggcttggtg cgcgaggctc 3660
 gcatttcgga tgttattgtc aataactttt tccttgccgg tggagaggtc ggccatcacc 3720
 gcccttctgt gatacctcgc ggcaatcctg atcagaacct cgggactcta caggcccttc 3780
 cgccgtcatg tcagatcagt gcttaccatc agttggctga ggaactaggc catcgcccgg 3840
 cccctgtcgc cgccgtcttg ccccttgcc ctgagcttga gcagggcctg ctctatatgc 3900
 cacaagaact tactgtgtcc gatagcgtgc tggtttttga gcttacggat atagtcact 3960
 gccgtatggc cgcaccaagc cagcgaaagg ctgttctctc aacgcttggtg gggaggtagc 4020
 gccgtaggac taaattatat gaggcggcgc attcagatgt ccgtgagtc ctagcgaggc 4080
 ttatccccac catcgggcct gttcgggcta ccacatgtga gctgtacgag ctggttgaag 4140
 ccatggtaga gaagggtcag gacggatctg ccgtcctaga gctcgacctt tgcaatcgtg 4200
 acgtctcgcg catcacattt ttccaaaagg attgcaataa gtttacaact ggtgagacta 4260
 tcgcccattg caaggttggc cagggcatac cggcctggag caagaccttc tgtgctctgt 4320
 ttggcccggtg gttccgcgcc attgaaaagg aaatattggc cctactcccg cctaatatct 4380
 tttatggcga cgcctatgag gagtcagtgt ttgctgccgc tgtgtccggg gcagggtcat 4440
 gtatggtatt tgaaaatgac ttctcagagt ttgacagtac ccagaataat ttctctctcg 4500
 gccttgagtg tgtggttatg gaggagtgcg gcatgcccc aagggttaatt aggttgtagc 4560
 atctgggtccg gtcagcctgg attttgcagg cgcggaagga gtctcttaag gggttttgga 4620
 agaagcactc tggtagcctt ggtacccttc tctggaacac tgtctggaac atggcgatta 4680
 tagcacattg ctaygagttc cgtgactttc gtgttgccgc cttcaagggt gatgattcag 4740
 tggctcctctg tagtgactac cgacagrgcc gtaacgcggc tgcccttaatt gcaggctgtg 4800
 ggctcaaatt gaaggttgat taccgcccta tcgggctata tgctggagtg gtggtggccc 4860
 cgggtttggg gacactgccc gatgtggtgc gttttgccgg tcggttatct gagaagaatt 4920
 ggggcccctgg cccggagcgt gctgagcagc tgcgtcttgc tgtttgtgat ttccctcgag 4980

gggttgacgaa tgttgcgag gtctgtgttg atgttggtgc cgtgtctat ggagttagcc 5040
 ccgggctggt acataacctt attggcatgc tgcagaccat tgctgatggc aaggccact 5100
 ttacagaraa tattaacct gtgcttgacc ttacaaattc catcatacaa cgggtggaat 5160
 gaataacatg tcttttgcac cgcctatggg atcaccatgc gccctagggc tgttctgttg 5220
 ttgctcttcg tgcttttgcc tatgtgtccc gcgcacccgg ccggccagcc gtctggccgc 5280
 cgtcgtgggc ggcgacggc cgggtgccgc ggtggtttct ggggtgacag ggttgattct 5340
 cagcccttcg cctcccta tttcatcca accaaccct tcgcgcgcga tgcgtttca 5400
 caaccgggg ctggaactcg cctcgacag ccgcgccgc ccttggytc cgttggcgt 5460
 gaccagtccc agcgccctc cgtgcccc cgtcgtgat ctgcccagc tggggctgcg 5520
 ccgctgactg ccgtgtcacc ggctcctgac acagccctg tacctgatgt tgactcacgt 5580
 ggtgctattc tgcgcggca gtacaatttg tccacgtccc cgtcacgtc atctgtcgt 5640
 tcgggtacta atttggtcct ctatgtgcc ccgtgaatc cctcttgcc tctccaggat 5700
 ggtaccaaca ctcatattat ggctactgag gcaccaatt atgcccagta tcgggtgtt 5760
 cgagctacaa tccgttatcg ccgctggtg ccgaatgcg ttggtggcta tgccatttc 5820
 atttctttct ggcccaaac tacaactacc cctactctg togatatgaa ttctattact 5880
 tccacygatg ttaggatttt ggttcagccc ggtattgct ccgagctagt catccccagt 5940
 gagcgcttc attaccgtaa tcaaggctgg cgtctgttg agaccacggg tgtggctgag 6000
 gaggaggcta cttcgggtc ggtaatgctt tgcattcatg gctctcctgt taattcctac 6060
 actaatacac cttacactgg tgcgtgggg cttcttgatt ttgcactaga gcttgaattt 6120
 aggaatttga caccgggaa caccaacacc cgtgtttccc ggtataccag cacagccgc 6180
 caccggctgc gccgtggtgc tgatgggact gctgagctta ctaccacagc agccacacgt 6240
 ttcatgaagg acctgcactt cgtggcacg aatggcgttg gtgaggtggg tcgtggtatc 6300
 gccctgacac tgttcaatct cgtgatacg cttctcggcg gtttaccgac agaattgatt 6360
 tcgtcggctg ggggccaact gttttactcc cgcgcggttg tctcagcaa tggcgagcca 6420
 acagtaaagt tatatacatc tgttgagaat gcgcagcaag acaaggcat caccattcca 6480
 catgatatag acctgggtga cttcgtgtg gttatccagg attatgataa ccagcaygag 6540
 caagaccgac ctactcgtc acctgcccc tctgcacct tctcagttct tcgtgccaat 6600
 gatgttttgt ggctttccct cactgccgt gagtatgacc agactacgta tgggtcgtcc 6660

accaacccta tgtatgtctc tgacacagtt acgcttggtta atgtggctac tgggtgtcag 6720
 gctgttgccc gctcccttga ttggtctaaa gttactctgg acggccgccc cttactacc 6780
 attcagcagt attctaagac attttatgtt ctcccgtcc gcgggaagct gtccttttg 6840
 gaggtctggca cgactaaggc cggtaccct tacaattata atactaccgc tagtgaccaa 6900
 attttgattg agaatgcggc cggccaccgt gtcgctatct ccacctatac cactagctta 6960
 ggtgcgggtc ctacctgat ctctgcggtc ggcgtactgg ctccacactc tgcccttgcc 7020
 gttcttgagg atactattga ttaccccgcc cgtgcccata cttttgatga tttttgccc 7080
 gagtgccgta ccctagggtt gcagggttgt gcattccagt ctactattgc tgagctccag 7140
 cgtttaaaaa tgaaggtagg taaaaccgg gagtcttaat taattccttc tgtgccccct 7200
 tcgtagtttc tttcgctttt atttcttatt tctgctttcc gcgctccctg gaaaaaaaaa 7260
 aaaaaaaaaa aaaaaaa 7277

<210> 165
 <211> 7277
 <212> DNA
 <213> Hepatitis E virus

<220>
 <221> CDS
 <222> (36)..(5162)
 <223> orf1

<220>
 <221> CDS
 <222> (5197)..(7179)
 <223> orf2

<220>
 <221> misc_feature
 <222> (5159)..(5527)
 <223> CDS- orf3

<220>
 <223> us2full

<400> 165
 tcgacagggg gcagaccacg tatgtggctg atgcc atg gag gcc cat cag ttc 53
 Met Glu Ala His Gln Phe
 1 5

att aag gct cct ggc att act act gct att gag cag gct gct ctg gct 101
 Ile Lys Ala Pro Gly Ile Thr Thr Ala Ile Glu Gln Ala Ala Leu Ala
 10 15 20

gcg gct aat tcc gcc ttg gcg aat gct gtg gtg gtt cgg ccg ttt ctt 149
 Ala Ala Asn Ser Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu

25	30	35	
tct cgt gtg caa act gag att ctt att aat ttg atg caa ccc cgg cag Ser Arg Val Gln Thr Glu Ile Leu Ile Asn Leu Met Gln Pro Arg Gln 40 45 50			197
ttg gtc ttc cgc cct gag gtg ctt tgg aat cat cct atc cag cgg gtt Leu Val Phe Arg Pro Glu Val Leu Trp Asn His Pro Ile Gln Arg Val 55 60 65 70			245
ata cat aat gaa tta gag cag tac tgc cgg gcc cgg gct ggt cgt tgt Ile His Asn Glu Leu Glu Gln Tyr Cys Arg Ala Arg Ala Gly Arg Cys 75 80 85			293
ttg gag gtt gga gcc cac ccg agg tcc att aat gac aac cct aat gtc Leu Glu Val Gly Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val 90 95 100			341
ttg cat agg tgt ttt ctt aga ccg gtc ggc cga gat gtt cag cgc tgg Leu His Arg Cys Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp 105 110 115			389
tat tct gcc cct acc cgt ggt cct gcg gcc aat tgc cgc cgc tcc gcg Tyr Ser Ala Pro Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala 120 125 130			437
ttg cgt ggt ctc ccc cct gtc gac cgc acc tat tgt ttt gat gga ttt Leu Arg Gly Leu Pro Pro Val Asp Arg Thr Tyr Cys Phe Asp Gly Phe 135 140 145 150			485
tcc cgt tgt gct ttt gct gca gag acc ggt gtg gcc ctt tac tct ttg Ser Arg Cys Ala Phe Ala Ala Glu Thr Gly Val Ala Leu Tyr Ser Leu 155 160 165			533
cat gac ctt tgg cca gct gat gtt gca gag gct atg gcc cgc cat ggg His Asp Leu Trp Pro Ala Asp Val Ala Glu Ala Met Ala Arg His Gly 170 175 180			581
atg aca cgc tta tac gcc gca ctg cac ctt ccc ccc gag gtg ctg cta Met Thr Arg Leu Tyr Ala Ala Leu His Leu Pro Pro Glu Val Leu Leu 185 190 195			629
cca ccc ggc acc tac cac aca acc tcg tac ctc ttg att cac gat ggc Pro Pro Gly Thr Tyr His Thr Thr Ser Tyr Leu Leu Ile His Asp Gly 200 205 210			677
aac cgc gct gtt gta act tac gag ggc gat act agt gcg ggc tat aat Asn Arg Ala Val Val Thr Tyr Glu Gly Asp Thr Ser Ala Gly Tyr Asn 215 220 225 230			725
cat gat gtc tcc ata ctt cgt gca tgg atc cgt act act aaa ata gtt His Asp Val Ser Ile Leu Arg Ala Trp Ile Arg Thr Thr Lys Ile Val 235 240 245			773
ggg gac cat cca ttg gtc ata gag cga gtg cgg gcc att ggg tgt cat Gly Asp His Pro Leu Val Ile Glu Arg Val Arg Ala Ile Gly Cys His 250 255 260			821

ttt	gtg	ctg	ctg	ctc	acc	gca	gcc	cct	gaa	ccg	tca	cct	atg	cct	tat	869
Phe	Val	Leu	Leu	Leu	Thr	Ala	Ala	Pro	Glu	Pro	Ser	Pro	Met	Pro	Tyr	
	265						270					275				
gtt	ccc	tac	cct	cgt	tca	acg	gag	gtg	tat	gtc	cgg	tct	ata	ttt	ggc	917
Val	Pro	Tyr	Pro	Arg	Ser	Thr	Glu	Val	Tyr	Val	Arg	Ser	Ile	Phe	Gly	
	280					285					290					
cct	ggc	ggc	tcc	cca	tcc	ttg	ttt	cca	tca	gcc	tgc	tct	act	aaa	tct	965
Pro	Gly	Gly	Ser	Pro	Ser	Leu	Phe	Pro	Ser	Ala	Cys	Ser	Thr	Lys	Ser	
295					300					305					310	
acc	ttt	cat	gct	gtc	ccg	gtt	cac	atc	tgg	gat	crg	ctc	atg	ctc	ttt	1013
Thr	Phe	His	Ala	Val	Pro	Val	His	Ile	Trp	Asp	Xaa	Leu	Met	Leu	Phe	
			315						320						325	
ggg	gcc	acc	ctg	rac	gat	cag	gcg	ttc	tgc	tgt	tca	cgg	ctt	atg	act	1061
Gly	Ala	Thr	Leu	Xaa	Asp	Gln	Ala	Phe	Cys	Cys	Ser	Arg	Leu	Met	Thr	
			330					335					340			
tac	ctc	cgt	ggg	att	agt	tat	aag	gtc	act	gtc	ggg	gcg	ctt	gtc	gct	1109
Tyr	Leu	Arg	Gly	Ile	Ser	Tyr	Lys	Val	Thr	Val	Gly	Ala	Leu	Val	Ala	
		345					350					355				
aat	gag	ggg	tgg	aac	gcc	tct	gag	gat	gct	ctt	act	gca	gtg	atc	act	1157
Asn	Glu	Gly	Trp	Asn	Ala	Ser	Glu	Asp	Ala	Leu	Thr	Ala	Val	Ile	Thr	
	360					365					370					
gcg	gcc	tat	ctg	acc	atc	tgc	cat	cag	cgt	tac	ctt	cgc	acc	cag	gcg	1205
Ala	Ala	Tyr	Leu	Thr	Ile	Cys	His	Gln	Arg	Tyr	Leu	Arg	Thr	Gln	Ala	
375					380					385					390	
att	tcc	aag	ggc	atg	cgc	cgg	ttg	gag	gtt	gag	cat	gct	cag	aaa	ttt	1253
Ile	Ser	Lys	Gly	Met	Arg	Arg	Leu	Glu	Val	Glu	His	Ala	Gln	Lys	Phe	
				395					400					405		
atc	aca	aga	ctc	tac	agc	tgg	cta	ttt	gag	aag	tct	ggc	cgt	gac	tac	1301
Ile	Thr	Arg	Leu	Tyr	Ser	Trp	Leu	Phe	Glu	Lys	Ser	Gly	Arg	Asp	Tyr	
			410					415					420			
atc	ccc	ggc	cgc	cag	ctt	caa	ttt	tat	gca	caa	tgc	cga	cgg	tgg	ctt	1349
Ile	Pro	Gly	Arg	Gln	Leu	Gln	Phe	Tyr	Ala	Gln	Cys	Arg	Arg	Trp	Leu	
		425					430					435				
tct	gca	ggc	ttc	cac	cta	rac	ccc	agg	rtg	ctt	gtc	ttt	gat	gaa	tca	1397
Ser	Ala	Gly	Phe	His	Leu	Xaa	Pro	Arg	Xaa	Leu	Val	Phe	Asp	Glu	Ser	
	440					445					450					
gtg	cca	tgc	cgt	tgc	agg	acg	ttt	ttg	aag	aag	gtc	gcg	ggg	aaa	ttc	1445
Val	Pro	Cys	Arg	Cys	Arg	Thr	Phe	Leu	Lys	Lys	Val	Ala	Gly	Lys	Phe	
455					460					465					470	
tgc	tgt	ttt	atg	cgg	tgg	ctg	ggg	cag	gag	tgt	acc	tgc	ttc	ttg	gag	1493
Cys	Cys	Phe	Met	Arg	Trp	Leu	Gly	Gln	Glu	Cys	Thr	Cys	Phe	Leu	Glu	
				475					480					485		

cca gcc gag ggt tta gtt ggt gat caa ggt cat gac aac gag gcc tat	1541
Pro Ala Glu Gly Leu Val Gly Asp Gln Gly His Asp Asn Glu Ala Tyr	
490 495 500	
gaa ggt tct gag gtc gac cca gct gag cct gca cat ctt gat gtc tcg	1589
Glu Gly Ser Glu Val Asp Pro Ala Glu Pro Ala His Leu Asp Val Ser	
505 510 515	
ggg act tat gcc gtc cat ggg cac cag ctt gag gcc ctc tat agg gca	1637
Gly Thr Tyr Ala Val His Gly His Gln Leu Glu Ala Leu Tyr Arg Ala	
520 525 530	
ctt aat gtc cca cat gat att gcc gct cga gcc tcc cga cta acg gct	1685
Leu Asn Val Pro His Asp Ile Ala Ala Arg Ala Ser Arg Leu Thr Ala	
535 540 545 550	
act gtt gag ctc gtt gct agt ccg gac cgc tta gag tgc cgc act gta	1733
Thr Val Glu Leu Val Ala Ser Pro Asp Arg Leu Glu Cys Arg Thr Val	
555 560 565	
ctt ggt aat aag acc ttc cgg acg acg gtg gtt gat ggc gcc cat ctt	1781
Leu Gly Asn Lys Thr Phe Arg Thr Thr Val Val Asp Gly Ala His Leu	
570 575 580	
gaa gcg aat ggc cct gag gag tat gtt ctg tca ttt gac gcc tct cgc	1829
Glu Ala Asn Gly Pro Glu Glu Tyr Val Leu Ser Phe Asp Ala Ser Arg	
585 590 595	
cag tct atg ggg gcc ggg tcg cac agc ctc act tat gag ctc acc cct	1877
Gln Ser Met Gly Ala Gly Ser His Ser Leu Thr Tyr Glu Leu Thr Pro	
600 605 610	
gcc ggt ctg cag gta aag att tca tct aat ggt ctg gat tgc act gcc	1925
Ala Gly Leu Gln Val Lys Ile Ser Ser Asn Gly Leu Asp Cys Thr Ala	
615 620 625 630	
aca ttc ccc ccy ggt ggc gcc cct agc gcc gcg ccg ggg gag gtg gcs	1973
Thr Phe Pro Xaa Gly Gly Ala Pro Ser Ala Ala Pro Gly Glu Val Xaa	
635 640 645	
gcc ttc tgc agt gct ctt tat aga tac aat agg ttc acc cag cgg cat	2021
Ala Phe Cys Ser Ala Leu Tyr Arg Tyr Asn Arg Phe Thr Gln Arg His	
650 655 660	
tcg ctg aca ggc gga cta tgg cta cat cct gag ggg ctg ctg ggt atc	2069
Ser Leu Thr Gly Gly Leu Trp Leu His Pro Glu Gly Leu Leu Gly Ile	
665 670 675	
ttc ccc cca ttc tcc cct ggg cat att tgg gag tct gct aac ccc ttt	2117
Phe Pro Pro Phe Ser Pro Gly His Ile Trp Glu Ser Ala Asn Pro Phe	
680 685 690	
tgc ggt gag ggg act ttg tat acc cga acc tgg tca acc tct ggt ttt	2165
Cys Gly Glu Gly Thr Leu Tyr Thr Arg Thr Trp Ser Thr Ser Gly Phe	
695 700 705 710	
tct agt gat ttc tcc ccc cct gag gcg gcc gct cct gct tcg gct gcc	2213

Ser	Ser	Asp	Phe	Ser	Pro	Pro	Glu	Ala	Ala	Ala	Pro	Ala	Ser	Ala	Ala		
				715					720					725			
gcc	ccg	ggg	ttg	ccc	tac	cct	act	cca	cct	gtt	agt	gat	atc	tgg	gtg	2261	
Ala	Pro	Gly	Leu	Pro	Tyr	Pro	Thr	Pro	Pro	Val	Ser	Asp	Ile	Trp	Val		
			730					735					740				
tta	cca	ccg	ccc	tca	gag	gaa	tct	cat	gtt	gat	gcg	gca	tct	gta	ccc	2309	
Leu	Pro	Pro	Pro	Ser	Glu	Glu	Ser	His	Val	Asp	Ala	Ala	Ser	Val	Pro		
			745				750					755					
tct	gtt	cct	gag	cct	gct	gga	ttg	acc	agc	cct	att	gtg	ctt	acc	ccc	2357	
Ser	Val	Pro	Glu	Pro	Ala	Gly	Leu	Thr	Ser	Pro	Ile	Val	Leu	Thr	Pro		
			760			765					770						
ccc	ccc	ccc	cct	cct	ccc	gtg	cgt	aag	ccg	gca	aca	tcc	ccg	cct	ccc	2405	
Pro	Pro	Pro	Pro	Pro	Val	Arg	Lys	Pro	Ala	Thr	Ser	Pro	Pro	Pro	Pro		
					780				785						790		
cgc	act	cgc	cgt	ctc	ctt	tac	acc	tac	ccc	gac	ggc	gcc	aag	gtg	tat	2453	
Arg	Thr	Arg	Arg	Leu	Leu	Tyr	Thr	Tyr	Pro	Asp	Gly	Ala	Lys	Val	Tyr		
				795					800					805			
gcg	ggg	tca	ttg	tkt	gag	tca	gac	tgt	gat	tgg	tta	gtc	aat	gcc	tca	2501	
Ala	Gly	Ser	Leu	Xaa	Glu	Ser	Asp	Cys	Asp	Trp	Leu	Val	Asn	Ala	Ser		
			810					815					820				
aac	cct	ggc	cat	cgc	ccc	ggg	ggt	ggc	ctc	tgc	cat	gct	ttt	tat	caa	2549	
Asn	Pro	Gly	His	Arg	Pro	Gly	Gly	Gly	Leu	Cys	His	Ala	Phe	Tyr	Gln		
			825				830					835					
cgt	ttc	cca	gaa	gcg	ttc	tac	tcg	act	gaa	ttc	atc	atg	cgc	gag	ggc	2597	
Arg	Phe	Pro	Glu	Ala	Phe	Tyr	Ser	Thr	Glu	Phe	Ile	Met	Arg	Glu	Gly		
			840			845					850						
ctt	gca	gca	tac	act	tta	acc	ccg	cgc	cct	att	atc	cat	gca	gtg	gct	2645	
Leu	Ala	Ala	Tyr	Thr	Leu	Thr	Pro	Arg	Pro	Ile	Ile	His	Ala	Val	Ala		
					860					865					870		
ccc	gac	tat	agg	gtt	gag	caa	aac	ccg	aag	agg	ctt	gag	gca	gcg	tac	2693	
Pro	Asp	Tyr	Arg	Val	Glu	Gln	Asn	Pro	Lys	Arg	Leu	Glu	Ala	Ala	Tyr		
				875					880					885			
cgg	gaa	act	tgc	tcc	cgt	cgt	ggc	acc	gct	gcc	tac	ccg	ctt	ttg	ggc	2741	
Arg	Glu	Thr	Cys	Ser	Arg	Arg	Gly	Thr	Ala	Ala	Tyr	Pro	Leu	Leu	Gly		
			890					895					900				
tcg	ggt	ata	tac	cag	gtc	cct	gtt	agc	ctc	agt	ttt	gat	gcc	tgg	gaa	2789	
Ser	Gly	Ile	Tyr	Gln	Val	Pro	Val	Ser	Leu	Ser	Phe	Asp	Ala	Trp	Glu		
			905				910					915					
cgc	aat	cac	cgc	ccc	ggc	gat	gag	ctt	tac	ttg	aca	gag	ccc	gcc	gca	2837	
Arg	Asn	His	Arg	Pro	Gly	Asp	Glu	Leu	Tyr	Leu	Thr	Glu	Pro	Ala	Ala		
			920			925					930						
gcc	tgg	ttt	gag	gct	aat	aag	ccg	gcg	cag	ccg	gcg	ctt	act	ata	act	2885	
Ala	Trp	Phe	Glu	Ala	Asn	Lys	Pro	Ala	Gln	Pro	Ala	Leu	Thr	Ile	Thr		

935	940	945	950	
gag gac acg gcc cgt acg gcc aac ctg gca tta gag att gat gcc gcc				2933
Glu Asp Thr Ala Arg Thr Ala Asn Leu Ala Leu Glu Ile Asp Ala Ala				
955		960	965	
aca gag gtt ggc cgt gct tgt gcc ggc tgc acc atc agc ccc ggg att				2981
Thr Glu Val Gly Arg Ala Cys Ala Gly Cys Thr Ile Ser Pro Gly Ile				
970		975	980	
gtg cac tat cag ttt acc gcc ggg gtc ccg ggc tca ggc aag tca agg				3029
Val His Tyr Gln Phe Thr Ala Gly Val Pro Gly Ser Gly Lys Ser Arg				
985		990	995	
tcc ata caa cag gga gat gtc gat gtg gtg gtt gtg ccc acc cgg gag				3077
Ser Ile Gln Gln Gly Asp Val Asp Val Val Val Pro Thr Arg Glu				
1000	1005	1010		
ctc cgt aac agc tgg cgt cgc cgg ggt ttt gcg gcc ttc aca cct cac				3125
Leu Arg Asn Ser Trp Arg Arg Arg Gly Phe Ala Ala Phe Thr Pro His				
1015	1020	1025	1030	
aca gcg gcc cgt gtt act atc ggc cgc cgc gtt gtg att gat gag gct				3173
Thr Ala Ala Arg Val Thr Ile Gly Arg Arg Val Val Ile Asp Glu Ala				
1035	1040	1045		
cca tct ctc cca ccg cac ctg ctg ctg tta cac atg cag cgg gcc tcc				3221
Pro Ser Leu Pro Pro His Leu Leu Leu Leu His Met Gln Arg Ala Ser				
1050	1055	1060		
tcg gtc cat ctc ctt ggt gat cca aac cag att cct gct att gat ttt				3269
Ser Val His Leu Leu Gly Asp Pro Asn Gln Ile Pro Ala Ile Asp Phe				
1065	1070	1075		
gag cat gcc ggc ctg gtc ccc gcg atc cgc ccc gag ctt gcg cca acg				3317
Glu His Ala Gly Leu Val Pro Ala Ile Arg Pro Glu Leu Ala Pro Thr				
1080	1085	1090		
agc tgg tgg cac gtt aca cac cgt tgc ccg gcc gat gtg tgc gag ctc				3365
Ser Trp Trp His Val Thr His Arg Cys Pro Ala Asp Val Cys Glu Leu				
1095	1100	1105	1110	
ata cgt ggg gcc tac ccc aaa att cag acc acg agc cgt gtg cta cgg				3413
Ile Arg Gly Ala Tyr Pro Lys Ile Gln Thr Thr Ser Arg Val Leu Arg				
1115	1120	1125		
tcc ctg ttt tgg aac gaa ccg gcc atc ggc caa aag ttg gtt ttt acg				3461
Ser Leu Phe Trp Asn Glu Pro Ala Ile Gly Gln Lys Leu Val Phe Thr				
1130	1135	1140		
cag gct gct aag gct gcc aac cct ggt gcg att acg gtt cac gaa gct				3509
Gln Ala Ala Lys Ala Ala Asn Pro Gly Ala Ile Thr Val His Glu Ala				
1145	1150	1155		
cag ggt gct act ttc acg gag acc aca att ata gcc acg gcc gac gct				3557
Gln Gly Ala Thr Phe Thr Glu Thr Thr Ile Ile Ala Thr Ala Asp Ala				
1160	1165	1170		

agg	ggc	ctc	att	cag	tca	tcc	cgg	gcc	cat	gct	ata	gtc	gca	ctc	acc	3605
Arg	Gly	Leu	Ile	Gln	Ser	Ser	Arg	Ala	His	Ala	Ile	Val	Ala	Leu	Thr	
1175				1180				1185				1190				
cgc	cat	act	gag	aag	tgt	gtt	att	ttg	gat	gcc	ccc	ggc	ttg	ttg	cgc	3653
Arg	His	Thr	Glu	Lys	Cys	Val	Ile	Leu	Asp	Ala	Pro	Gly	Leu	Leu	Arg	
1195				1200				1205								
gag	gtc	ggc	att	tcg	gat	gtt	att	gtc	aat	aac	ttt	ttc	ctt	gcc	ggc	3701
Glu	Val	Gly	Ile	Ser	Asp	Val	Ile	Val	Asn	Asn	Phe	Phe	Leu	Ala	Gly	
1210				1215				1220								
gga	gag	gtc	ggc	cat	cac	cgc	cct	tct	gtg	ata	cct	cgc	ggc	aat	cct	3749
Gly	Glu	Val	Gly	His	His	Arg	Pro	Ser	Val	Ile	Pro	Arg	Gly	Asn	Pro	
1225				1230				1235								
gat	cag	aac	ctc	ggg	act	cta	cag	gcc	ttt	ccg	ccg	tca	tgt	cag	atc	3797
Asp	Gln	Asn	Leu	Gly	Thr	Leu	Gln	Ala	Phe	Pro	Pro	Ser	Cys	Gln	Ile	
1240				1245				1250								
agt	gct	tac	cat	cag	ttg	gct	gag	gaa	cta	ggc	cat	cgc	ccg	gcc	cct	3845
Ser	Ala	Tyr	His	Gln	Leu	Ala	Glu	Glu	Leu	Gly	His	Arg	Pro	Ala	Pro	
1255				1260				1265				1270				
gtc	gcc	gcc	gtc	ttg	ccc	cct	tgc	cct	gag	ctt	gag	cag	ggc	ctg	ctc	3893
Val	Ala	Ala	Val	Leu	Pro	Pro	Cys	Pro	Glu	Leu	Glu	Gln	Gly	Leu	Leu	
1275				1280				1285								
tat	atg	cca	caa	gaa	ctt	act	gtg	tcc	gat	agc	gtg	ctg	gtt	ttt	gag	3941
Tyr	Met	Pro	Gln	Glu	Leu	Thr	Val	Ser	Asp	Ser	Val	Leu	Val	Phe	Glu	
1290				1295				1300								
ctt	acg	gat	ata	gtc	cac	tgc	cgt	atg	gcc	gcc	cca	agc	cag	cga	aag	3989
Leu	Thr	Asp	Ile	Val	His	Cys	Arg	Met	Ala	Ala	Pro	Ser	Gln	Arg	Lys	
1305				1310				1315								
gct	gtt	ctc	tca	acg	ctt	gtg	ggg	agg	tac	ggc	cgt	agg	act	aaa	tta	4037
Ala	Val	Leu	Ser	Thr	Leu	Val	Gly	Arg	Tyr	Gly	Arg	Arg	Thr	Lys	Leu	
1320				1325				1330								
tat	gag	gcg	gcg	cat	tca	gat	gtc	cgt	gag	tcc	cta	gcg	agg	ttt	atc	4085
Tyr	Glu	Ala	Ala	His	Ser	Asp	Val	Arg	Glu	Ser	Leu	Ala	Arg	Phe	Ile	
1335				1340				1345				1350				
ccc	acc	atc	ggg	cct	gtt	cgg	gct	acc	aca	tgt	gag	ctg	tac	gag	ctg	4133
Pro	Thr	Ile	Gly	Pro	Val	Arg	Ala	Thr	Thr	Cys	Glu	Leu	Tyr	Glu	Leu	
1355				1360				1365								
gtt	gaa	gcc	atg	gta	gag	aag	ggc	cag	gac	gga	tct	gcc	gtc	cta	gag	4181
Val	Glu	Ala	Met	Val	Glu	Lys	Gly	Gln	Asp	Gly	Ser	Ala	Val	Leu	Glu	
1370				1375				1380								
ctc	gac	ctt	tgc	aat	cgt	gac	gtc	tcg	cgc	atc	aca	ttt	ttc	caa	aag	4229
Leu	Asp	Leu	Cys	Asn	Arg	Asp	Val	Ser	Arg	Ile	Thr	Phe	Phe	Gln	Lys	
1385				1390				1395								

gat tgc aat aag ttt aca act ggt gag act atc gcc cat ggc aag gtt Asp Cys Asn Lys Phe Thr Thr Gly Glu Thr Ile Ala His Gly Lys Val 1400 1405 1410	4277
ggc cag ggc ata tgc gcc tgg agc aag acc ttc tgt gct ctg ttt ggc Gly Gln Gly Ile Ser Ala Trp Ser Lys Thr Phe Cys Ala Leu Phe Gly 1415 1420 1425 1430	4325
ccg tgg ttc cgc gcc att gaa aag gaa ata ttg gcc cta ctc ccg cct Pro Trp Phe Arg Ala Ile Glu Lys Glu Ile Leu Ala Leu Leu Pro Pro 1435 1440 1445	4373
aat atc ttt tat ggc gac gcc tat gag gag tca gtg ttt gct gcc gct Asn Ile Phe Tyr Gly Asp Ala Tyr Glu Glu Ser Val Phe Ala Ala Ala 1450 1455 1460	4421
gtg tcc ggg gca ggg tca tgt atg gta ttt gaa aat gac ttc tca gag Val Ser Gly Ala Gly Ser Cys Met Val Phe Glu Asn Asp Phe Ser Glu 1465 1470 1475	4469
ttt gac agt acc cag aat aat ttc tct ctc ggc ctt gag tgt gtg gtt Phe Asp Ser Thr Gln Asn Asn Phe Ser Leu Gly Leu Glu Cys Val Val 1480 1485 1490	4517
atg gag gag tgc ggc atg ccc caa tgg tta att agg ttg tac cat ctg Met Glu Glu Cys Gly Met Pro Gln Trp Leu Ile Arg Leu Tyr His Leu 1495 1500 1505 1510	4565
gtc cgg tca gcc tgg att ttg cag gcg ccg aag gag tct ctt aag ggg Val Arg Ser Ala Trp Ile Leu Gln Ala Pro Lys Glu Ser Leu Lys Gly 1515 1520 1525	4613
ttt tgg aag aag cac tct ggt gag cct ggt acc ctt ctc tgg aac act Phe Trp Lys Lys His Ser Gly Glu Pro Gly Thr Leu Leu Trp Asn Thr 1530 1535 1540	4661
gtc tgg aac atg gcg att ata gca cat tgc tay gag ttc cgt gac ttt Val Trp Asn Met Ala Ile Ile Ala His Cys Xaa Glu Phe Arg Asp Phe 1545 1550 1555	4709
cgt gtt gcc gcc ttc aag ggt gat gat tca gtg gtc ctc tgt agt gac Arg Val Ala Ala Phe Lys Gly Asp Asp Ser Val Val Leu Cys Ser Asp 1560 1565 1570	4757
tac cga cag rgc cgt aac gcg gct gcc tta att gca ggc tgt ggg ctc Tyr Arg Gln Xaa Arg Asn Ala Ala Ala Leu Ile Ala Gly Cys Gly Leu 1575 1580 1585 1590	4805
aaa ttg aag gtt gat tac cgc cct atc ggg cta tat gct gga gtg gtg Lys Leu Lys Val Asp Tyr Arg Pro Ile Gly Leu Tyr Ala Gly Val Val 1595 1600 1605	4853
gtg gcc ccc ggt ttg ggg aca ctg ccc gat gtg gtg cgt ttt gcc ggt Val Ala Pro Gly Leu Gly Thr Leu Pro Asp Val Val Arg Phe Ala Gly 1610 1615 1620	4901
cgg tta tct gag aag aat tgg ggc cct ggc ccg gag cgt gct gag cag	4949

Arg Leu Ser Glu Lys Asn Trp Gly Pro Gly Pro Glu Arg Ala Glu Gln	
1625 1630 1635	
ctg cgt ctt gct gtt tgt gat ttc ctt cga ggg ttg acg aat gtt gcg	4997
Leu Arg Leu Ala Val Cys Asp Phe Leu Arg Gly Leu Thr Asn Val Ala	
1640 1645 1650	
cag gtc tgt gtt gat gtt gtg tcc cgt gtc tat gga gtt agc ccc ggg	5045
Gln Val Cys Val Asp Val Val Ser Arg Val Tyr Gly Val Ser Pro Gly	
1655 1660 1665 1670	
ctg gta cat aac ctt att ggc atg ctg cag acc att gct gat ggc aag	5093
Leu Val His Asn Leu Ile Gly Met Leu Gln Thr Ile Ala Asp Gly Lys	
1675 1680 1685	
gcc cac ttt aca gar aat att aaa cct gtg ctt gac ctt aca aat tcc	5141
Ala His Phe Thr Xaa Asn Ile Lys Pro Val Leu Asp Leu Thr Asn Ser	
1690 1695 1700	
atc ata caa cgg gtg gaa tga ataacatgtc ttttgcacg cccatgggat cacc	5196
Ile Ile Gln Arg Val Glu	
1705	
atg cgc cct agg gct gtt ctg ttg ttg ctc ttc gtg ctt ttg cct atg	5244
Met Arg Pro Arg Ala Val Leu Leu Leu Leu Phe Val Leu Leu Pro Met	
1710 1715 1720 1725	
ctg ccc gcg cca ccg gcc ggc cag ccg tct ggc cgc cgt cgt ggg cgg	5292
Leu Pro Ala Pro Pro Ala Gly Gln Pro Ser Gly Arg Arg Arg Gly Arg	
1730 1735 1740	
cgc agc ggc ggt gcc ggc ggt ggt ttc tgg ggt gac agg gtt gat tct	5340
Arg Ser Gly Gly Ala Gly Gly Gly Phe Trp Gly Asp Arg Val Asp Ser	
1745 1750 1755	
cag ccc ttc gcc ctc ccc tat att cat cca acc aac ccc ttc gcc gcc	5388
Gln Pro Phe Ala Leu Pro Tyr Ile His Pro Thr Asn Pro Phe Ala Ala	
1760 1765 1770	
gat gtc gtt tca caa ccc ggg gct gga act cgc cct cga cag ccg ccc	5436
Asp Val Val Ser Gln Pro Gly Ala Gly Thr Arg Pro Arg Gln Pro Pro	
1775 1780 1785	
cgc ccc ctt ggy tcc gct tgg cgt gac cag tcc cag cgc ccc tcc gct	5484
Arg Pro Leu Xaa Ser Ala Trp Arg Asp Gln Ser Gln Arg Pro Ser Ala	
1790 1795 1800 1805	
gcc ccc cgt cgt cga tct gcc cca gct ggg gct gcg ccg ctg act gcc	5532
Ala Pro Arg Arg Arg Ser Ala Pro Ala Gly Ala Ala Pro Leu Thr Ala	
1810 1815 1820	
gtg tca ccg gct cct gac aca gcc cct gta cct gat gtt gac tca cgt	5580
Val Ser Pro Ala Pro Asp Thr Ala Pro Val Pro Asp Val Asp Ser Arg	
1825 1830 1835	
ggt gct att ctg cgc cgg cag tac aat ttg tcc acg tcc ccg ctc acg	5628
Gly Ala Ile Leu Arg Arg Gln Tyr Asn Leu Ser Thr Ser Pro Leu Thr	

1840	1845	1850	
tca tct gtc gct tcg ggt act aat ttg gtc ctc tat gct gcc ccg ctg			5676
Ser Ser Val Ala Ser Gly Thr Asn Leu Val Leu Tyr Ala Ala Pro Leu			
1855	1860	1865	
aat ccc ctc ttg cct ctc cag gat ggt acc aac act cat att atg gct			5724
Asn Pro Leu Leu Pro Leu Gln Asp Gly Thr Asn Thr His Ile Met Ala			
1870	1875	1880	1885
act gag gca tcc aat tat gcc cag tat cgg gtt gtt cga gct aca atc			5772
Thr Glu Ala Ser Asn Tyr Ala Gln Tyr Arg Val Val Arg Ala Thr Ile			
1890	1895	1900	
cgt tat cgc ccg ctg gtg ccg aat gcc gtt ggt ggc tat gcc att tcc			5820
Arg Tyr Arg Pro Leu Val Pro Asn Ala Val Gly Gly Tyr Ala Ile Ser			
1905	1910	1915	
att tct ttc tgg ccc caa act aca act acc cct act tct gtc gat atg			5868
Ile Ser Phe Trp Pro Gln Thr Thr Thr Thr Pro Thr Ser Val Asp Met			
1920	1925	1930	
aat tct att act tcc acy gat gtt agg att ttg gtt cag ccc ggt att			5916
Asn Ser Ile Thr Ser Xaa Asp Val Arg Ile Leu Val Gln Pro Gly Ile			
1935	1940	1945	
gcc tcc gag cta gtc atc ccc agt gag cgc ctt cat tac cgt aat caa			5964
Ala Ser Glu Leu Val Ile Pro Ser Glu Arg Leu His Tyr Arg Asn Gln			
1950	1955	1960	1965
ggc tgg cgc tct gtt gag acc acg ggt gtg gct gag gag gag gct act			6012
Gly Trp Arg Ser Val Glu Thr Thr Gly Val Ala Glu Glu Glu Ala Thr			
1970	1975	1980	
tcc ggt ctg gta atg ctt tgc att cat ggc tct cct gtt aat tcc tac			6060
Ser Gly Leu Val Met Leu Cys Ile His Gly Ser Pro Val Asn Ser Tyr			
1985	1990	1995	
act aat aca cct tac act ggt gcg ctg ggg ctt ctt gat ttt gca cta			6108
Thr Asn Thr Pro Tyr Thr Gly Ala Leu Gly Leu Leu Asp Phe Ala Leu			
2000	2005	2010	
gag ctt gaa ttt agg aat ttg aca ccc ggg aac acc aac acc cgt gtt			6156
Glu Leu Glu Phe Arg Asn Leu Thr Pro Gly Asn Thr Asn Thr Arg Val			
2015	2020	2025	
tcc cgg tat acc agc aca gcc cgc cac cgg ctg cgc cgt ggt gct gat			6204
Ser Arg Tyr Thr Ser Thr Ala Arg His Arg Leu Arg Arg Gly Ala Asp			
2030	2035	2040	2045
ggg act gct gag ctt act acc aca gca gcc aca cgt ttc atg aag gac			6252
Gly Thr Ala Glu Leu Thr Thr Ala Ala Thr Arg Phe Met Lys Asp			
2050	2055	2060	
ctg cac ttc gct ggc acg aat ggc gtt ggt gag gtg ggt cgt ggt atc			6300
Leu His Phe Ala Gly Thr Asn Gly Val Gly Glu Val Gly Arg Gly Ile			
2065	2070	2075	

gcc ctg aca ctg ttc aat ctc gct gat acg ctt ctc ggc ggt tta ccg Ala Leu Thr Leu Phe Asn Leu Ala Asp Thr Leu Leu Gly Gly Leu Pro 2080 2085 2090	6348
aca gaa ttg att tcg tcg gct ggg ggc caa ctg ttt tac tcc cgc ccg Thr Glu Leu Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro 2095 2100 2105	6396
gtt gtc tca gcc aat ggc gag cca aca gta aag tta tat aca tct gtt Val Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val 2110 2115 2120 2125	6444
gag aat gcg cag caa gac aag ggc atc acc att cca cat gat ata gac Glu Asn Ala Gln Gln Asp Lys Gly Ile Thr Ile Pro His Asp Ile Asp 2130 2135 2140	6492
ctg ggt gac tcc cgt gtg gtt atc cag gat tat gat aac cag cay gag Leu Gly Asp Ser Arg Val Val Ile Gln Asp Tyr Asp Asn Gln Xaa Glu 2145 2150 2155	6540
caa gac cga cct act ccg tca cct gcc ccc tct cgc ccc ttc tca gtt Gln Asp Arg Pro Thr Pro Ser Pro Ala Pro Ser Arg Pro Phe Ser Val 2160 2165 2170	6588
ctt cgt gcc aat gat gtt ttg tgg ctt tcc ctc act gcc gct gag tat Leu Arg Ala Asn Asp Val Leu Trp Leu Ser Leu Thr Ala Ala Glu Tyr 2175 2180 2185	6636
gac cag act acg tat ggg tcg tcc acc aac cct atg tat gtc tct gac Asp Gln Thr Thr Tyr Gly Ser Ser Thr Asn Pro Met Tyr Val Ser Asp 2190 2195 2200 2205	6684
aca gtt acg ctt gtt aat gtg gct act ggt gct cag gct gtt gcc cgc Thr Val Thr Leu Val Asn Val Ala Thr Gly Ala Gln Ala Val Ala Arg 2210 2215 2220	6732
tcc ctt gat tgg tct aaa gtt act ctg gac ggc cgc ccc ctt act acc Ser Leu Asp Trp Ser Lys Val Thr Leu Asp Gly Arg Pro Leu Thr Thr 2225 2230 2235	6780
att cag cag tat tct aag aca ttt tat gtt ctc ccg ctc cgc ggg aag Ile Gln Gln Tyr Ser Lys Thr Phe Tyr Val Leu Pro Leu Arg Gly Lys 2240 2245 2250	6828
ctg tcc ttt tgg gag gct ggc acg act aag gcc ggc tac cct tac aat Leu Ser Phe Trp Glu Ala Gly Thr Thr Lys Ala Gly Tyr Pro Tyr Asn 2255 2260 2265	6876
tat aat act acc gct agt gac caa att ttg att gag aat gcg gcc ggc Tyr Asn Thr Thr Ala Ser Asp Gln Ile Leu Ile Glu Asn Ala Ala Gly 2270 2275 2280 2285	6924
cac cgt gtc gct att tcc acc tat acc act agc tta ggt gcc ggt cct His Arg Val Ala Ile Ser Thr Tyr Thr Thr Ser Leu Gly Ala Gly Pro 2290 2295 2300	6972

acc tcg atc tct gcg gtc ggc gta ctg gct cca cac tct gcc ett gcc 7020
 Thr Ser Ile Ser Ala Val Gly Val Leu Ala Pro His Ser Ala Leu Ala
 2305 2310 2315

gtt ctt gag gat act att gat tac ccc gcc cgt gcc cat act ttt gat 7068
 Val Leu Glu Asp Thr Ile Asp Tyr Pro Ala Arg Ala His Thr Phe Asp
 2320 2325 2330

gat ttt tgc ccg gag tgc cgt acc cta ggt ttg cag ggt tgt gca ttc 7116
 Asp Phe Cys Pro Glu Cys Arg Thr Leu Gly Leu Gln Gly Cys Ala Phe
 2335 2340 2345

cag tct act att gct gag ctg cag cgt tta aaa atg aag gta ggt aaa 7164
 Gln Ser Thr Ile Ala Glu Leu Gln Arg Leu Lys Met Lys Val Gly Lys
 2350 2355 2360 2365

acc cgg gag tct taa ttaattcctt ctgtgcccc ttctgtagttt ctttcgcttt 7219
 Thr Arg Glu Ser
 2370

tatttcttat ttctgctttc cgcgctccct ggaaaaaaaa aaaaaaaaaa aaaaaaaaaa 7277

<210> 166

<211> 1708

<212> PRT

<213> Hepatitis E virus

<400> 166

Met Glu Ala His Gln Phe Ile Lys Ala Pro Gly Ile Thr Thr Ala Ile
 1 5 10 15

Glu Gln Ala Ala Leu Ala Ala Ala Asn Ser Ala Leu Ala Asn Ala Val
 20 25 30

Val Val Arg Pro Phe Leu Ser Arg Val Gln Thr Glu Ile Leu Ile Asn
 35 40 45

Leu Met Gln Pro Arg Gln Leu Val Phe Arg Pro Glu Val Leu Trp Asn
 50 55 60

His Pro Ile Gln Arg Val Ile His Asn Glu Leu Glu Gln Tyr Cys Arg
 65 70 75 80

Ala Arg Ala Gly Arg Cys Leu Glu Val Gly Ala His Pro Arg Ser Ile
 85 90 95

Asn Asp Asn Pro Asn Val Leu His Arg Cys Phe Leu Arg Pro Val Gly
 100 105 110

Arg Asp Val Gln Arg Trp Tyr Ser Ala Pro Thr Arg Gly Pro Ala Ala
 115 120 125

Asn Cys Arg Arg Ser Ala Leu Arg Gly Leu Pro Pro Val Asp Arg Thr
 130 135 140

Tyr Cys Phe Asp Gly Phe Ser Arg Cys Ala Phe Ala Ala Glu Thr Gly

145		150		155		160
Val Ala Leu Tyr Ser	Leu His Asp Leu Trp Pro Ala Asp Val Ala Glu					
	165			170		175
Ala Met Ala Arg His Gly Met Thr Arg Leu Tyr Ala Ala Leu His Leu						
	180			185		190
Pro Pro Glu Val Leu Leu Pro Pro Gly Thr Tyr His Thr Thr Ser Tyr						
	195			200		205
Leu Leu Ile His Asp Gly Asn Arg Ala Val Val Thr Tyr Glu Gly Asp						
	210			215		220
Thr Ser Ala Gly Tyr Asn His Asp Val Ser Ile Leu Arg Ala Trp Ile						
	225			230		235
Arg Thr Thr Lys Ile Val Gly Asp His Pro Leu Val Ile Glu Arg Val						
	245			250		255
Arg Ala Ile Gly Cys His Phe Val Leu Leu Leu Thr Ala Ala Pro Glu						
	260			265		270
Pro Ser Pro Met Pro Tyr Val Pro Tyr Pro Arg Ser Thr Glu Val Tyr						
	275			280		285
Val Arg Ser Ile Phe Gly Pro Gly Gly Ser Pro Ser Leu Phe Pro Ser						
	290			295		300
Ala Cys Ser Thr Lys Ser Thr Phe His Ala Val Pro Val His Ile Trp						
	305			310		315
Asp Xaa Leu Met Leu Phe Gly Ala Thr Leu Xaa Asp Gln Ala Phe Cys						
	325			330		335
Cys Ser Arg Leu Met Thr Tyr Leu Arg Gly Ile Ser Tyr Lys Val Thr						
	340			345		350
Val Gly Ala Leu Val Ala Asn Glu Gly Trp Asn Ala Ser Glu Asp Ala						
	355			360		365
Leu Thr Ala Val Ile Thr Ala Ala Tyr Leu Thr Ile Cys His Gln Arg						
	370			375		380
Tyr Leu Arg Thr Gln Ala Ile Ser Lys Gly Met Arg Arg Leu Glu Val						
	385			390		395
Glu His Ala Gln Lys Phe Ile Thr Arg Leu Tyr Ser Trp Leu Phe Glu						
	405			410		415
Lys Ser Gly Arg Asp Tyr Ile Pro Gly Arg Gln Leu Gln Phe Tyr Ala						
	420			425		430
Gln Cys Arg Arg Trp Leu Ser Ala Gly Phe His Leu Xaa Pro Arg Xaa						
	435			440		445
Leu Val Phe Asp Glu Ser Val Pro Cys Arg Cys Arg Thr Phe Leu Lys						

450				455				460							
Lys	Val	Ala	Gly	Lys	Phe	Cys	Cys	Phe	Met	Arg	Trp	Leu	Gly	Gln	Glu
465					470					475					480
Cys	Thr	Cys	Phe	Leu	Glu	Pro	Ala	Glu	Gly	Leu	Val	Gly	Asp	Gln	Gly
				485					490					495	
His	Asp	Asn	Glu	Ala	Tyr	Glu	Gly	Ser	Glu	Val	Asp	Pro	Ala	Glu	Pro
			500					505					510		
Ala	His	Leu	Asp	Val	Ser	Gly	Thr	Tyr	Ala	Val	His	Gly	His	Gln	Leu
		515					520					525			
Glu	Ala	Leu	Tyr	Arg	Ala	Leu	Asn	Val	Pro	His	Asp	Ile	Ala	Ala	Arg
		530				535					540				
Ala	Ser	Arg	Leu	Thr	Ala	Thr	Val	Glu	Leu	Val	Ala	Ser	Pro	Asp	Arg
545					550					555					560
Leu	Glu	Cys	Arg	Thr	Val	Leu	Gly	Asn	Lys	Thr	Phe	Arg	Thr	Thr	Val
				565					570					575	
Val	Asp	Gly	Ala	His	Leu	Glu	Ala	Asn	Gly	Pro	Glu	Glu	Tyr	Val	Leu
			580					585					590		
Ser	Phe	Asp	Ala	Ser	Arg	Gln	Ser	Met	Gly	Ala	Gly	Ser	His	Ser	Leu
		595					600					605			
Thr	Tyr	Glu	Leu	Thr	Pro	Ala	Gly	Leu	Gln	Val	Lys	Ile	Ser	Ser	Asn
		610				615					620				
Gly	Leu	Asp	Cys	Thr	Ala	Thr	Phe	Pro	Xaa	Gly	Gly	Ala	Pro	Ser	Ala
625					630					635					640
Ala	Pro	Gly	Glu	Val	Xaa	Ala	Phe	Cys	Ser	Ala	Leu	Tyr	Arg	Tyr	Asn
			645						650					655	
Arg	Phe	Thr	Gln	Arg	His	Ser	Leu	Thr	Gly	Gly	Leu	Trp	Leu	His	Pro
			660					665					670		
Glu	Gly	Leu	Leu	Gly	Ile	Phe	Pro	Pro	Phe	Ser	Pro	Gly	His	Ile	Trp
		675					680					685			
Glu	Ser	Ala	Asn	Pro	Phe	Cys	Gly	Glu	Gly	Thr	Leu	Tyr	Thr	Arg	Thr
		690				695					700				
Trp	Ser	Thr	Ser	Gly	Phe	Ser	Ser	Asp	Phe	Ser	Pro	Pro	Glu	Ala	Ala
705					710					715					720
Ala	Pro	Ala	Ser	Ala	Ala	Ala	Pro	Gly	Leu	Pro	Tyr	Pro	Thr	Pro	Pro
				725					730					735	
Val	Ser	Asp	Ile	Trp	Val	Leu	Pro	Pro	Pro	Ser	Glu	Glu	Ser	His	Val
			740					745					750		
Asp	Ala	Ala	Ser	Val	Pro	Ser	Val	Pro	Glu	Pro	Ala	Gly	Leu	Thr	Ser

755					760					765					
Pro	Ile	Val	Leu	Thr	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Val	Arg	Lys	Pro
770					775					780					
Ala	Thr	Ser	Pro	Pro	Pro	Arg	Thr	Arg	Arg	Leu	Leu	Tyr	Thr	Tyr	Pro
785					790					795					800
Asp	Gly	Ala	Lys	Val	Tyr	Ala	Gly	Ser	Leu	Xaa	Glu	Ser	Asp	Cys	Asp
				805					810					815	
Trp	Leu	Val	Asn	Ala	Ser	Asn	Pro	Gly	His	Arg	Pro	Gly	Gly	Gly	Leu
			820					825					830		
Cys	His	Ala	Phe	Tyr	Gln	Arg	Phe	Pro	Glu	Ala	Phe	Tyr	Ser	Thr	Glu
		835					840					845			
Phe	Ile	Met	Arg	Glu	Gly	Leu	Ala	Ala	Tyr	Thr	Leu	Thr	Pro	Arg	Pro
		850				855					860				
Ile	Ile	His	Ala	Val	Ala	Pro	Asp	Tyr	Arg	Val	Glu	Gln	Asn	Pro	Lys
865					870					875					880
Arg	Leu	Glu	Ala	Ala	Tyr	Arg	Glu	Thr	Cys	Ser	Arg	Arg	Gly	Thr	Ala
				885					890					895	
Ala	Tyr	Pro	Leu	Leu	Gly	Ser	Gly	Ile	Tyr	Gln	Val	Pro	Val	Ser	Leu
			900					905					910		
Ser	Phe	Asp	Ala	Trp	Glu	Arg	Asn	His	Arg	Pro	Gly	Asp	Glu	Leu	Tyr
		915					920					925			
Leu	Thr	Glu	Pro	Ala	Ala	Ala	Trp	Phe	Glu	Ala	Asn	Lys	Pro	Ala	Gln
		930				935					940				
Pro	Ala	Leu	Thr	Ile	Thr	Glu	Asp	Thr	Ala	Arg	Thr	Ala	Asn	Leu	Ala
945					950					955					960
Leu	Glu	Ile	Asp	Ala	Ala	Thr	Glu	Val	Gly	Arg	Ala	Cys	Ala	Gly	Cys
				965					970					975	
Thr	Ile	Ser	Pro	Gly	Ile	Val	His	Tyr	Gln	Phe	Thr	Ala	Gly	Val	Pro
			980					985					990		
Gly	Ser	Gly	Lys	Ser	Arg	Ser	Ile	Gln	Gln	Gly	Asp	Val	Asp	Val	Val
		995				1000					1005				
Val	Val	Pro	Thr	Arg	Glu	Leu	Arg	Asn	Ser	Trp	Arg	Arg	Arg	Gly	Phe
					1010			1015			1020				
Ala	Ala	Phe	Thr	Pro	His	Thr	Ala	Ala	Arg	Val	Thr	Ile	Gly	Arg	Arg
025					1030					1035					1040
Val	Val	Ile	Asp	Glu	Ala	Pro	Ser	Leu	Pro	Pro	His	Leu	Leu	Leu	Leu
				1045				1050					1055		
His	Met	Gln	Arg	Ala	Ser	Ser	Val	His	Leu	Leu	Gly	Asp	Pro	Asn	Gln

1060	1065	1070
Ile Pro Ala Ile Asp Phe Glu His Ala Gly Leu Val Pro Ala Ile Arg 1075	1080	1085
Pro Glu Leu Ala Pro Thr Ser Trp Trp His Val Thr His Arg Cys Pro 1090	1095	1100
Ala Asp Val Cys Glu Leu Ile Arg Gly Ala Tyr Pro Lys Ile Gln Thr 105	1110	1115 1120
Thr Ser Arg Val Leu Arg Ser Leu Phe Trp Asn Glu Pro Ala Ile Gly 1125	1130	1135
Gln Lys Leu Val Phe Thr Gln Ala Ala Lys Ala Ala Asn Pro Gly Ala 1140	1145	1150
Ile Thr Val His Glu Ala Gln Gly Ala Thr Phe Thr Glu Thr Thr Ile 1155	1160	1165
Ile Ala Thr Ala Asp Ala Arg Gly Leu Ile Gln Ser Ser Arg Ala His 1170	1175	1180
Ala Ile Val Ala Leu Thr Arg His Thr Glu Lys Cys Val Ile Leu Asp 185	1190	1195 1200
Ala Pro Gly Leu Leu Arg Glu Val Gly Ile Ser Asp Val Ile Val Asn 1205	1210	1215
Asn Phe Phe Leu Ala Gly Gly Glu Val Gly His His Arg Pro Ser Val 1220	1225	1230
Ile Pro Arg Gly Asn Pro Asp Gln Asn Leu Gly Thr Leu Gln Ala Phe 1235	1240	1245
Pro Pro Ser Cys Gln Ile Ser Ala Tyr His Gln Leu Ala Glu Glu Leu 1250	1255	1260
Gly His Arg Pro Ala Pro Val Ala Ala Val Leu Pro Pro Cys Pro Glu 265	1270	1275 1280
Leu Glu Gln Gly Leu Leu Tyr Met Pro Gln Glu Leu Thr Val Ser Asp 1285	1290	1295
Ser Val Leu Val Phe Glu Leu Thr Asp Ile Val His Cys Arg Met Ala 1300	1305	1310
Ala Pro Ser Gln Arg Lys Ala Val Leu Ser Thr Leu Val Gly Arg Tyr 1315	1320	1325
Gly Arg Arg Thr Lys Leu Tyr Glu Ala Ala His Ser Asp Val Arg Glu 1330	1335	1340
Ser Leu Ala Arg Phe Ile Pro Thr Ile Gly Pro Val Arg Ala Thr Thr 345	1350	1355 1360
Cys Glu Leu Tyr Glu Leu Val Glu Ala Met Val Glu Lys Gly Gln Asp		

1365										1370										1375																			
Gly Ser Ala Val Leu Glu Leu Asp Leu Cys Asn Arg Asp Val Ser Arg																																							
1380										1385										1390																			
Ile Thr Phe Phe Gln Lys Asp Cys Asn Lys Phe Thr Thr Gly Glu Thr																																							
1395										1400										1405																			
Ile Ala His Gly Lys Val Gly Gln Gly Ile Ser Ala Trp Ser Lys Thr																																							
1410										1415										1420																			
Phe Cys Ala Leu Phe Gly Pro Trp Phe Arg Ala Ile Glu Lys Glu Ile																																							
425										1430										1435										1440									
Leu Ala Leu Leu Pro Pro Asn Ile Phe Tyr Gly Asp Ala Tyr Glu Glu																																							
1445										1450										1455																			
Ser Val Phe Ala Ala Ala Val Ser Gly Ala Gly Ser Cys Met Val Phe																																							
1460										1465										1470																			
Glu Asn Asp Phe Ser Glu Phe Asp Ser Thr Gln Asn Asn Phe Ser Leu																																							
1475										1480										1485																			
Gly Leu Glu Cys Val Val Met Glu Glu Cys Gly Met Pro Gln Trp Leu																																							
1490										1495										1500																			
Ile Arg Leu Tyr His Leu Val Arg Ser Ala Trp Ile Leu Gln Ala Pro																																							
505										1510										1515										1520									
Lys Glu Ser Leu Lys Gly Phe Trp Lys Lys His Ser Gly Glu Pro Gly																																							
1525										1530										1535																			
Thr Leu Leu Trp Asn Thr Val Trp Asn Met Ala Ile Ile Ala His Cys																																							
1540										1545										1550																			
Xaa Glu Phe Arg Asp Phe Arg Val Ala Ala Phe Lys Gly Asp Asp Ser																																							
1555										1560										1565																			
Val Val Leu Cys Ser Asp Tyr Arg Gln Xaa Arg Asn Ala Ala Ala Leu																																							
1570										1575										1580																			
Ile Ala Gly Cys Gly Leu Lys Leu Lys Val Asp Tyr Arg Pro Ile Gly																																							
585										1590										1595										1600									
Leu Tyr Ala Gly Val Val Val Ala Pro Gly Leu Gly Thr Leu Pro Asp																																							
1605										1610										1615																			
Val Val Arg Phe Ala Gly Arg Leu Ser Glu Lys Asn Trp Gly Pro Gly																																							
1620										1625										1630																			
Pro Glu Arg Ala Glu Gln Leu Arg Leu Ala Val Cys Asp Phe Leu Arg																																							
1635										1640										1645																			
Gly Leu Thr Asn Val Ala Gln Val Cys Val Asp Val Val Ser Arg Val																																							
1650										1655										1660																			
Tyr Gly Val Ser Pro Gly Leu Val His Asn Leu Ile Gly Met Leu Gln																																							

665	1670								1675				1680			
Thr	Ile	Ala	Asp	Gly	Lys	Ala	His	Phe	Thr	Xaa	Asn	Ile	Lys	Pro	Val	
				1685					1690					1695		
Leu	Asp	Leu	Thr	Asn	Ser	Ile	Ile	Gln	Arg	Val	Glu					
				1700					1705							
<210> 167																
<211> 660																
<212> PRT																
<213> Hepatitis E virus																
<400> 167																
Met	Arg	Pro	Arg	Ala	Val	Leu	Leu	Leu	Leu	Phe	Val	Leu	Leu	Pro	Met	
1				5				10				15				
Leu	Pro	Ala	Pro	Pro	Ala	Gly	Gln	Pro	Ser	Gly	Arg	Arg	Arg	Gly	Arg	
			20				25				30					
Arg	Ser	Gly	Gly	Ala	Gly	Gly	Gly	Phe	Trp	Gly	Asp	Arg	Val	Asp	Ser	
		35			40			45								
Gln	Pro	Phe	Ala	Leu	Pro	Tyr	Ile	His	Pro	Thr	Asn	Pro	Phe	Ala	Ala	
		50			55			60								
Asp	Val	Val	Ser	Gln	Pro	Gly	Ala	Gly	Thr	Arg	Pro	Arg	Gln	Pro	Pro	
65				70				75				80				
Arg	Pro	Leu	Xaa	Ser	Ala	Trp	Arg	Asp	Gln	Ser	Gln	Arg	Pro	Ser	Ala	
			85				90				95					
Ala	Pro	Arg	Arg	Arg	Ser	Ala	Pro	Ala	Gly	Ala	Ala	Pro	Leu	Thr	Ala	
			100				105				110					
Val	Ser	Pro	Ala	Pro	Asp	Thr	Ala	Pro	Val	Pro	Asp	Val	Asp	Ser	Arg	
		115			120			125								
Gly	Ala	Ile	Leu	Arg	Arg	Gln	Tyr	Asn	Leu	Ser	Thr	Ser	Pro	Leu	Thr	
		130			135			140								
Ser	Ser	Val	Ala	Ser	Gly	Thr	Asn	Leu	Val	Leu	Tyr	Ala	Ala	Pro	Leu	
145				150				155				160				
Asn	Pro	Leu	Leu	Pro	Leu	Gln	Asp	Gly	Thr	Asn	Thr	His	Ile	Met	Ala	
			165				170				175					
Thr	Glu	Ala	Ser	Asn	Tyr	Ala	Gln	Tyr	Arg	Val	Val	Arg	Ala	Thr	Ile	
			180				185				190					
Arg	Tyr	Arg	Pro	Leu	Val	Pro	Asn	Ala	Val	Gly	Gly	Tyr	Ala	Ile	Ser	
		195			200			205								
Ile	Ser	Phe	Trp	Pro	Gln	Thr	Thr	Thr	Thr	Pro	Thr	Ser	Val	Asp	Met	
		210			215			220								

Asn	Ser	Ile	Thr	Ser	Xaa	Asp	Val	Arg	Ile	Leu	Val	Gln	Pro	Gly	Ile	
225					230					235					240	
Ala	Ser	Glu	Leu	Val	Ile	Pro	Ser	Glu	Arg	Leu	His	Tyr	Arg	Asn	Gln	
				245					250					255		
Gly	Trp	Arg	Ser	Val	Glu	Thr	Thr	Gly	Val	Ala	Glu	Glu	Glu	Ala	Thr	
			260					265						270		
Ser	Gly	Leu	Val	Met	Leu	Cys	Ile	His	Gly	Ser	Pro	Val	Asn	Ser	Tyr	
		275					280						285			
Thr	Asn	Thr	Pro	Tyr	Thr	Gly	Ala	Leu	Gly	Leu	Leu	Asp	Phe	Ala	Leu	
		290				295					300					
Glu	Leu	Glu	Phe	Arg	Asn	Leu	Thr	Pro	Gly	Asn	Thr	Asn	Thr	Arg	Val	
305					310					315					320	
Ser	Arg	Tyr	Thr	Ser	Thr	Ala	Arg	His	Arg	Leu	Arg	Arg	Gly	Ala	Asp	
				325					330					335		
Gly	Thr	Ala	Glu	Leu	Thr	Thr	Thr	Ala	Ala	Thr	Arg	Phe	Met	Lys	Asp	
			340					345					350			
Leu	His	Phe	Ala	Gly	Thr	Asn	Gly	Val	Gly	Glu	Val	Gly	Arg	Gly	Ile	
		355					360					365				
Ala	Leu	Thr	Leu	Phe	Asn	Leu	Ala	Asp	Thr	Leu	Leu	Gly	Gly	Leu	Pro	
		370				375					380					
Thr	Glu	Leu	Ile	Ser	Ser	Ala	Gly	Gly	Gln	Leu	Phe	Tyr	Ser	Arg	Pro	
385					390					395					400	
Val	Val	Ser	Ala	Asn	Gly	Glu	Pro	Thr	Val	Lys	Leu	Tyr	Thr	Ser	Val	
				405					410					415		
Glu	Asn	Ala	Gln	Gln	Asp	Lys	Gly	Ile	Thr	Ile	Pro	His	Asp	Ile	Asp	
			420					425					430			
Leu	Gly	Asp	Ser	Arg	Val	Val	Ile	Gln	Asp	Tyr	Asp	Asn	Gln	Xaa	Glu	
		435					440					445				
Gln	Asp	Arg	Pro	Thr	Pro	Ser	Pro	Ala	Pro	Ser	Arg	Pro	Phe	Ser	Val	
		450				455					460					
Leu	Arg	Ala	Asn	Asp	Val	Leu	Trp	Leu	Ser	Leu	Thr	Ala	Ala	Glu	Tyr	
465					470					475				480		
Asp	Gln	Thr	Thr	Tyr	Gly	Ser	Ser	Thr	Asn	Pro	Met	Tyr	Val	Ser	Asp	
				485					490					495		
Thr	Val	Thr	Leu	Val	Asn	Val	Ala	Thr	Gly	Ala	Gln	Ala	Val	Ala	Arg	
			500					505					510			
Ser	Leu	Asp	Trp	Ser	Lys	Val	Thr	Leu	Asp	Gly	Arg	Pro	Leu	Thr	Thr	
		515					520					525				

Ile Gln Gln Tyr Ser Lys Thr Phe Tyr Val Leu Pro Leu Arg Gly Lys
530 535 540

Leu Ser Phe Trp Glu Ala Gly Thr Thr Lys Ala Gly Tyr Pro Tyr Asn
545 550 555 560

Tyr Asn Thr Thr Ala Ser Asp Gln Ile Leu Ile Glu Asn Ala Ala Gly
565 570 575

His Arg Val Ala Ile Ser Thr Tyr Thr Thr Ser Leu Gly Ala Gly Pro
580 585 590

Thr Ser Ile Ser Ala Val Gly Val Leu Ala Pro His Ser Ala Leu Ala
595 600 605

Val Leu Glu Asp Thr Ile Asp Tyr Pro Ala Arg Ala His Thr Phe Asp
610 615 620

Asp Phe Cys Pro Glu Cys Arg Thr Leu Gly Leu Gln Gly Cys Ala Phe
625 630 635 640

Gln Ser Thr Ile Ala Glu Leu Gln Arg Leu Lys Met Lys Val Gly Lys
645 650 655

Thr Arg Glu Ser
660

<210> 168

<211> 122

<212> PRT

<213> Hepatitis E virus

<220>

<223> us2 orf3

<400> 168

Met Asn Asn Met Ser Phe Ala Ser Pro Met Gly Ser Pro Cys Ala Leu
1 5 10 15

Gly Leu Phe Cys Cys Cys Ser Ser Cys Phe Cys Leu Cys Cys Pro Arg
20 25 30

His Arg Pro Ala Ser Arg Leu Ala Ala Val Val Gly Gly Ala Ala Ala
35 40 45

Val Pro Ala Val Val Ser Gly Val Thr Gly Leu Ile Leu Ser Pro Ser
50 55 60

Pro Ser Pro Ile Phe Ile Gln Pro Thr Pro Ser Pro Pro Met Ser Phe
65 70 75 80

His Asn Pro Gly Leu Glu Leu Ala Leu Asp Ser Arg Pro Ala Pro Leu
85 90 95

Xaa Pro Leu Gly Val Thr Ser Pro Ser Ala Pro Pro Leu Pro Pro Val
100 105 110

Val Asp Leu Pro Gln Leu Gly Leu Arg Arg
 115 120

<210> 169
 <211> 33
 <212> PRT
 <213> Hepatitis E virus

<220>
 <223> M 4-2

<400> 169
 Ala Asn Gln Pro Gly His Leu Ala Pro Leu Gly Glu Ile Arg Pro Ser
 1 5 10 15

Ala Pro Pro Leu Pro Pro Val Ala Asp Leu Pro Gln Pro Gly Leu Arg
 20 25 30

Arg

<210> 170
 <211> 48
 <212> PRT
 <213> Hepatitis E virus

<220>
 <223> M 3-2e

<400> 170
 Thr Phe Asp Tyr Pro Gly Arg Ala His Thr Phe Asp Asp Phe Cys Pro
 1 5 10 15

Glu Cys Arg Ala Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr Val
 20 25 30

Ala Glu Leu Gln Arg Leu Lys Val Lys Val Gly Lys Thr Arg Glu Leu
 35 40 45

<210> 171
 <211> 33
 <212> PRT
 <213> Hepatitis E virus

<220>
 <223> B 4-2

<400> 171
 Ala Asn Pro Pro Asp His Ser Ala Pro Leu Gly Val Thr Arg Pro Ser
 1 5 10 15

Ala Pro Pro Leu Pro His Val Val Asp Leu Pro Gln Leu Gly Pro Arg
 20 25 30

Arg

<210> 172
 <211> 48
 <212> PRT
 <213> Hepatitis E virus

<220>
 <223> B 3-2e

<400> 172
 Thr Leu Asp Tyr Pro Ala Arg Ala His Thr Phe Asp Asp Phe Cys Pro
 1 5 10 15
 Glu Cys Arg Pro Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr Val
 20 25 30
 Ala Glu Leu Gln Arg Leu Lys Met Lys Val Gly Lys Thr Arg Glu Leu
 35 40 45

<210> 173
 <211> 33
 <212> PRT
 <213> Hepatitis E virus

<220>
 <223> ORF3 (u4.2)

<400> 173
 Asp Ser Arg Pro Ala Pro Ser Val Pro Leu Gly Val Thr Ser Pro Ser
 1 5 10 15
 Ala Pro Pro Leu Pro Pro Val Val Asp Leu Pro Gln Leu Gly Leu Arg
 20 25 30

Arg

<210> 174
 <211> 48
 <212> PRT
 <213> Hepatitis E virus

<220>
 <223> ORF2 (u3.2e)

<400> 174

Thr Val Asp Tyr Pro Ala Arg Ala His Thr Phe Asp Asp Phe Cys Pro
 1 5 10 15

Glu Cys Arg Thr Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr Ile
 20 25 30

Ala Glu Leu Gln Arg Leu Lys Met Lys Val Gly Lys Thr Arg Glu Ser
 35 40 45

<210> 175

<211> 33

<212> PRT

<213> Hepatitis E virus

<220>

<223> US 4-2

<400> 175

Asp Ser Arg Pro Ala Pro Ser Val Pro Leu Gly Val Thr Ser Pro Ser
 1 5 10 15

Ala Pro Pro Leu Pro Pro Val Val Asp Leu Pro Gln Leu Gly Leu Arg
 20 25 30

Cys

<210> 176

<211> 48

<212> PRT

<213> Hepatitis E virus

<220>

<223> US 3-2e

<400> 176

Thr Val Asp Tyr Pro Ala Arg Ala His Thr Phe Asp Asp Phe Cys Pro
 1 5 10 15

Glu Cys Arg Thr Leu Gly Val Gln Gly Cys Ala Phe Gln Ser Thr Ile
 20 25 30

Ala Glu Val Gln Arg Leu Lys Met Lys Val Gly Lys Thr Arg Glu Val
 35 40 45

<210> 177

<211> 21

<212> DNA
 <213> Hepatitis E virus

<220>

<223> HEVConsORF1-s2

<400> 177

ctgccytkgc gaatgctgtg g 21

<210> 178

<211> 24

<212> DNA

<213> Hepatitis E virus

<220>

<223> HEVConsORF1-a2

<400> 178

ggcagwrtac carcgtgaa catc 24

<210> 179

<211> 294

<212> DNA

<213> Hepatitis E virus

<220>

<223> z12-orf1 (G.S.)

<400> 179

tggcattact actgccattg agcaagctgc tctggctgcg gccaatctg ccttggcgaa 60

tgctgtggtg gttcggccgt ttttatctcg ttacagact gagattctta ttaatttgat 120

gcaaccccg cagttggtct ttgacctga ggtgttctgg aaccatccca tccaacgtgt 180

tatacataat gaattggagc agtactgccg ggcccgggcc ggtcgctgtc tggaaattgg 240

agcccatcca aggtcaatca atgataatcc taatgttctg catcggtgtt tcct 294

<210> 180

<211> 418

<212> DNA

<213> Hepatitis E virus

<220>

<223> z12-orf1.con

<400> 180

ctggcattac tactgctatt gagcaagctg ctctgggtgc ggccaattct gccttggcga 60

atgctgtggt ggttcggccg tttttatctc gtttacagac tgagattctt attaatttga 120

tgcaaccccg acagttggtc tttegacctg aggtgttctg gaaccatccc atccaacgtg 180

ttatacataa tgaattggag cagtactgcc gggcccgggc cggtcgctgt ctggaaattg 240
 gagcccatcc aaggtcaatc aatgataatc ctaatgttct gcatcgggtgc tttttacgac 300
 cggtcggggag ggacgttcag cgctgggtact ccgccccccac ccgtggcccc gcggccaact 360
 gccgccggtc tgcgctgcgt ggtctccccc ctgtcgaccg cacttactgc ctcgatgg 418

<210> 181
 <211> 197
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> z12-orf2.con

<400> 181
 gacagaatta atttcgtcgg ctgggggtca actgttctac tcccgccttg tcgtctcagc 60
 caatggcgag ccgactgtca agttatacac atctgttgag aatgcacagc aggataaggg 120
 gatagctatt ccacatgaca tagatttggg cgactctcgt ttggtaatcc aggattatga 180
 taaccaacac gaacaag 197

<210> 182
 <211> 25
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> HEVConsORF2/3-s1

<400> 182
 gtatcggkyk gaatgaataa catgt 25

<210> 183
 <211> 25
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> HEVConsORF2/3-a1

<400> 183
 aggggttggg tggatgaata taggg 25

<210> 184
 <211> 234
 <212> DNA
 <213> Hepatitis E virus

<220>

<223> z12-orf23.con

<400> 184

gtatcggtt gaatgaataa catgttttgc gcatcgccca tgggatcacc atgcgccta 60
 ggggtgttct gttgttgttc ctctgttttc tgcctatgct gcccgcgcca ccggccggcc 120
 agycgactgg ccgcccgtct gggcggcgca gcggcggtgc cggcggtggt ttctgggggtg 180
 acagggttga ttctcagccc ttccgctccc cctatatcca tccaaccaac ccc 234

<210> 185

<211> 890

<212> DNA

<213> Hepatitis E virus

<220>

<223> z12-3p.race

<400> 185

gtcgtctcgg ccaatggcga gccgactgtc aagttataca catctgttga gaatgcacag 60
 caggataagg ggatagctat tccacatgac atagatttgg gcgactctcg tttggtaatc 120
 caggattacg ataatcagca cgagcaggac cggcccaccc ctccgcccgc cccgtctcgt 180
 cctttctcgg tcttcgcgc taatgatgct ttgtggcttt ctcttaccgc tgctgagtat 240
 gaccagacta catatgggtc gtccaccaac ccgatgtatg tctcagacac tgttacattt 300
 gtcaatgtgg ccacaggggc tcaggctgtc gcccgttctc ttgattgggc taaagttacc 360
 ctggacggcc gccctcttac taccatccag cagtactcta agacatttta tgttctccca 420
 ctccgcccga agttatcttt ttgggaggct ggcacaacta aagccgggta cccttataat 480
 tataacacaa ctgctagtga ccagattctg attgaaaacg cggctggcca tctgtctcgt 540
 atatctactt atactactag cctgggcgcc ggccctgtgt cagtttctgc ggttggtgtg 600
 ttagccccac actcgagcct tgctattctt gaagacactg ttgactatcc ggcccgtgct 660
 cacacttttg atgacttctg tccggaatgc cgtgcctgg gtctgcaggg gtgtgctttt 720
 caatctacta tcgctgagct ccagcgtctt aaaatgaagg taggcaaac ccgggagttt 780
 taattaattc ttcttgtgcc ccttcacgg ttctcgtttt atttctttct tctgcctccc 840
 gcgctccctg gaaaaaaaaa aaaaaaaaaa gtactagtcg acgcgtggcc 890

<210> 186

<211> 919

<212> DNA

<213> Hepatitis E virus

<220>

<223> z12-3p.con

<400> 186

gacagaatta atttcgtcgg ctgggggtca actgtttctac tcccgccttg tcgtctcagc 60
 caatggcgag ccgactgtca agttatacac atctgttgag aatgcacagc aggataaggg 120
 gatagctatt ccacatgaca tagatttggg cgactctcgt ttggtaatcc aggattacga 180
 taatcagcac gagcaggacc ggcccccccc ttgcgccgcc ccgtctcgtc ctttctcggg 240
 cctccgcgct aatgatgctt tgtggctttc tcttaccgct gctgagtatg accagactac 300
 atatgggtcg tccaccaacc cgatgtatgt ctacagacact gttacatttg tcaatgtggc 360
 cacaggggct caggctgtcg cccgttctct tgattgggtct aaagttaccg tggacggccg 420
 ccctcttact accatccagc agtactctaa gacattttat gttctccacac ttgcggggaa 480
 gttatctttt tgggaggctg gcacaactaa agccgggttac ccttataatt ataacacaac 540
 tgctagtga cagattctga ttgaaaacgc ggctggccat cgtgtcgtc taactactta 600
 tactactagc ctgggcgcgc gccctgtgtc agtttctgcg gttggtgtgt tagccccaca 660
 ctcgagcctt gctattcttg aagacactgt tgactatccg gcccggtgct acacttttga 720
 tgacttctgt ccggaatgcc gtgccttggg totgcagggg tgtgcttttc aatctactat 780
 cgctgagctc cagcgtctta aaatgaaggt aggcaaaacc cgggagtttt aattaattct 840
 tcttgtgcc ccttcacggt tctcgttcta tttctttctt ctgcctcccg cgctccctgg 900
 aaaaaaaaaa aaaaaaaaaa 919

<210> 187

<211> 138

<212> PRT

<213> Hepatitis E virus

<220>

<223> z12-orf1.pep

<400> 187

Gly Ile Thr Thr Ala Ile Glu Gln Ala Ala Leu Gly Ala Ala Asn Ser
 1 5 10 15
 Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu Ser Arg Leu Gln
 20 25 30
 Thr Glu Ile Leu Ile Asn Leu Met Gln Pro Arg Gln Leu Val Phe Arg
 35 40 45
 Pro Glu Val Phe Trp Asn His Pro Ile Gln Arg Val Ile His Asn Glu
 50 55 60

Leu Glu Gln Tyr Cys Arg Ala Arg Ala Gly Arg Cys Leu Glu Ile Gly
65 70 75 80

Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Leu His Arg Cys
85 90 95

Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp Tyr Ser Ala Pro
100 105 110

Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu Arg Gly Leu
115 120 125

Pro Pro Val Asp Arg Thr Tyr Cys Leu Asp
130 135

<210> 188

<211> 61

<212> PRT

<213> Hepatitis E virus

<220>

<223> z12-orf2-5'.pep

<400> 188

Met Arg Pro Arg Val Val Leu Leu Leu Phe Leu Val Phe Leu Pro Met
1 5 10 15

Leu Pro Ala Pro Pro Ala Gly Gln Xaa Thr Gly Arg Arg Arg Gly Arg
20 25 30

Arg Ser Gly Gly Ala Gly Gly Gly Phe Trp Gly Asp Arg Val Asp Ser
35 40 45

Gln Pro Phe Ala Leu Pro Tyr Ile His Pro Thr Asn Pro
50 55 60

<210> 189

<211> 276

<212> PRT

<213> Hepatitis E virus

<220>

<223> z12-orf2-3'.pep

<400> 189

Thr Glu Leu Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro
1 5 10 15

Val Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val
20 25 30

Glu Asn Ala Gln Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp
35 40 45

Leu Gly Asp Ser Arg Leu Val Ile Gln Asp Tyr Asp Asn Gln His Glu
 50 55 60
 Gln Asp Arg Pro Thr Pro Ser Pro Ala Pro Ser Arg Pro Phe Ser Val
 65 70 75 80
 Leu Arg Ala Asn Asp Ala Leu Trp Leu Ser Leu Thr Ala Ala Glu Tyr
 85 90 95
 Asp Gln Thr Thr Tyr Gly Ser Ser Thr Asn Pro Met Tyr Val Ser Asp
 100 105 110
 Thr Val Thr Phe Val Asn Val Ala Thr Gly Ala Gln Ala Val Ala Arg
 115 120 125
 Ser Leu Asp Trp Ser Lys Val Thr Leu Asp Gly Arg Pro Leu Thr Thr
 130 135 140
 Ile Gln Gln Tyr Ser Lys Thr Phe Tyr Val Leu Pro Leu Arg Gly Lys
 145 150 155 160
 Leu Ser Phe Trp Glu Ala Gly Thr Thr Lys Ala Gly Tyr Pro Tyr Asn
 165 170 175
 Tyr Asn Thr Thr Ala Ser Asp Gln Ile Leu Ile Glu Asn Ala Ala Gly
 180 185 190
 His Arg Val Ala Ile Ser Thr Tyr Thr Thr Ser Leu Gly Ala Gly Pro
 195 200 205
 Val Ser Val Ser Ala Val Gly Val Leu Ala Pro His Ser Ser Leu Ala
 210 215 220
 Ile Leu Glu Asp Thr Val Asp Tyr Pro Ala Arg Ala His Thr Phe Asp
 225 230 235 240
 Asp Phe Cys Pro Glu Cys Arg Ala Leu Gly Leu Gln Gly Cys Ala Phe
 245 250 255
 Gln Ser Thr Ile Ala Glu Leu Gln Arg Leu Lys Met Lys Val Gly Lys
 260 265 270
 Thr Arg Glu Phe
 275

<210> 190

<211> 74

<212> PRT

<213> Hepatitis E virus

<220>

<223> z12-orf3.pep

<400> 190

Met Asn Asn Met Phe Cys Ala Ser Pro Met Gly Ser Pro Cys Ala Leu
 1 5 10 15

Gly Leu Phe Cys Cys Cys Ser Ser Cys Phe Cys Leu Cys Cys Pro Arg
 20 25 30
 His Arg Pro Ala Ser Arg Leu Ala Ala Val Val Gly Gly Ala Ala Ala
 35 40 45
 Val Pro Ala Val Val Ser Gly Val Thr Gly Leu Ile Leu Ser Pro Ser
 50 55 60
 Pro Ser Pro Ile Phe Ile Gln Pro Thr Pro
 65 70

<210> 191
 <211> 408
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> pJOorf3-29.seq

<400> 191
 gaattcatga ataacatgtc ttttgcacgc cccatgggat caccatgcgc cctagggctg 60
 ttctgttggt gctcttcgtg cttttgccta tgcctgccgc gccaccggcc agccagccgt 120
 ctggcgcgcg tcgtgggcgg cgcagcggcg gtgcggcg tggtttctgg ggtgacaggg 180
 ttgattctca gcccttcgcc ctccctata ttcattcaac caacccttc gccgccgatg 240
 tcgtttcaca acccggggct ggaactcgcc ctgcacagcc gcccgcgcc cttggctccg 300
 cttggcgtga ccagtccag cgccttcg ctgcccccg tcgtgatct gccccagctt 360
 ggtctgcgcc gcgactacaa ggacgacgat gacaagtaat aaggatcc 408

<210> 192
 <211> 1026
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> cksorf2m-2.seq

<400> 192
 gaattcatgg gtgctgatgg gactgctgag cttactacca cagcagccac acgtttcatg 60
 aaggacctgc atttcgctgg cacgaatggc gttggtgagg tgggtcgtgg tategcctg 120
 aactgttca atctcgctga tacgcttctc ggcgggttac cgacagaatt gatttcgtcg 180
 gctggggggc aactgtttta ctccgcgcc gttgtctcag ccaatggcga gccaacagta 240
 aagttatata catctgttga gaatgcgcag caagacaagg gcattcaccat tccacatgat 300

```

atagacctgg gtgactcccg tgtgggtatc caggattatg ataaccagca tgagcaagac 360
cgacctactc cgtcacctgc cccctctcgc cccttctcag ttcttcgtgc caatgatgtt 420
ttgtggcttt cctcactgc cgctgagtat gaccagacta cgtatgggtc gtccaccaac 480
cctatgtatg tctctgacac agttacgctt gttaatgtgg ctactgggtc tcaggctgtt 540
gcccgcctcc ttgattggtc taaagttact ctggacggcc gcccccttac taccattcag 600
cagtattcta agacatttta tgttctcccg ctcccgggga agctgtcctt ttgggaggct 660
ggcacgacta aggccggcta cccttacaat tataatacta ccgctagtga ccaaattttg 720
attgagaatg cggccggcca ccgtgtcgtt atttccacct ataccactag cttagggtgc 780
ggctctacct cgatctctgc ggtcggcgta ctggctccac actctgccct tgccgttctt 840
gaggatacta ttgattaccc cgcccggtgc catacttttg atgatttttg cccggagtgc 900
cgtaccctag gtttgcaggg ttgtgcattc cagtctacta ttgctgagct ccagcgttta 960
aaaatgaagg taggtaaaac ccgggagtct gactacaagg acgacgatga caagtaataa 1020
ggatcc 1026

```

<210> 193

<211> 1389

<212> DNA

<213> Hepatitis E virus

<220>

<223> CKSORF32M-3.seq

<400> 193

```

gaattcatga ataacatgtc ttttgcacgc cccatgggat caccatgcgc cctagggctg 60
ttctgttggt gctcttcgtg cttttgccta tgctgccgcg gccaccggcc agccagccgt 120
ctggccgcgc tcgtgggcgg cgtagcggcg gtgccggcgg tggtttctgg ggtgacaggg 180
ttgattctca gcccttcgcc ctccctata ttcattcaac caaccccttc gccgccgatg 240
tcgtttcaca acccggggct ggaactcgcc ctgcacagcc gccccgccc cttggctccg 300
cttggcgtga ccagtccag cgccctccg ctgcccccg tcgtcgatct gccccagctt 360
ggctctgcgc gcggtgctga tgggactgct gagcttacta ccacagcagc cacacgtttc 420
atgaaggacc tgcacttcgc tggcacgaat ggcgttggtg aggtgggtcg tggtatcgcc 480
ctgacactgt tcaatctcgc tgatacgctt ctggcggtt taccgacaga attgatttcg 540
tcggctgggg gccaaactgt ttactccgcg ccggttgtct cagccaatgg cgagccaaca 600
gtaaagttat atacatctgt tgagaatgcg cagcaagaca agggcatcac cattccacat 660

```

gatatagacc tgggtgactc ccgtgtggtt atccaggatt atgataacca gcatgagcaa 720
gaccgaccta ctccgtcacc tgccccctct cgcctctctc cagttcttcg tgccaatgat 780
gttttgtggc tttccctcac tgccgctgag tatgaccaga ctacgtatgg gtcgtccacc 840
aaccctatgt atgtctctga cacagttacg cttgttaatg tggctactgg tgctcaggct 900
gttgcccgct ccttgattg gtctaaagtt actctggacg gccgccccct tactaccatt 960
cagcagtatt ctaagacatt ttatgttctc ccgtcccg ggaagctgtc cttttgggag 1020
gctggcacga ctaaggccg ctacccttac aattataata ctaccgctag tgaccaaatt 1080
ttgattgaga atgcggccg ccaccgtgtc gctatttcca cctataccac tagcttaggt 1140
gccggtccta cctcgatctc tgccgtcggc gtactggctc cacactctgc ccttgccgtt 1200
cttgaggata ctattgatta ccccgccgt gcccatactt ttgatgattt ttgccggag 1260
tgccgtaccc taggtttgca gggttgtgca ttccagtcta ctattgctga gctccagcgt 1320
ttaaaaatga aggtaggtaa aaccggggag tctgactaca aggacgacga tgacaagtaa 1380
taaggatcc 1389

<210> 194
<211> 408
<212> DNA
<213> Hepatitis E virus

<220>
<223> plorf3-12.con

<400> 194
gaattcatga ataacatgtc ttttgcacg cccatgggat caccatgcgc cctagggctg 60
ttctgttggt gctcttcgtg cttttgccta tgctgcccgc gccaccggcc ggccagccgt 120
ctggccgcgc tcgtgggcgg cgcagcggcg gtgcggcggt tggtttctgg ggtgacaggg 180
ttgattctca gcccttcgcc ctccccata ttcatccaac caacccttc gccgccgatg 240
tcgtttcaca acccgggggt ggaactcgcc ctgcacagcc gcccgcccc cttggctccg 300
cttggcgtga ccagtcacg cgcctctccg ctgccccccg tcgtcgatct gccccagctt 360
ggtctgcgcc gcgactacaa ggacgacgat gacaagtaat aaggatcc 408

<210> 195
<211> 1026
<212> DNA
<213> Hepatitis E virus

<220>

<223> plorf2.2-6.seq

<400> 195

```

gaattcatgg gtgctgatgg gactgctgag cttactacca cagcagccac acgtttcatg 60
aaggacctgc acttcgctgg cacgaatggc gttggtgagg tgggtcgtgg tatcgccctg 120
acactgttca atctcgctga tacgcttctc ggcggtttac cgacagaatt gatttcgctg 180
gctggggggcc aactgtttta ctcccgcccg gttgtctcag ccaatggcga gccaacagta 240
aagttatata catctgttga gaatgcgcag caagacaagg gcatcaccat tccacatgat 300
atagacctgg gtgactcccc tgtggttatc caggattatg ataaccagca tgagcaagac 360
cgacctactc cgtcacctgc cccctctcgc cccttctcag ttcttcgtgc caatgatgtt 420
ttgtggcttt ccctcactgc cgctgagtat gaccagacta cgtatgggtc gtccaccaac 480
cctatgtatg tctctgacac agttacgctt gttaatgtgg ctactggtgc tcaggctgtt 540
gcccgcctcc ttgattggtc taaagttact ctggacggcc gcccccttac taccattcag 600
cagtattcta agacatttta tgttctcccg ctccgcggga agctgtcctt ttgggaggct 660
ggcacgacta aggccggcta cccttacaat tataatacta ccgctagtga ccaaattttg 720
attgagaatg cggccggcca ccgtgtcgtc atttccacct ataccactag cttagggtgcc 780
ggtcctacct cgatctctgc ggtcggcgta ctggctccac actctgccct tgccgttctt 840
gaggatacta ttgattaccc cgcccggtgc catacttttg atgatttttg cccggagtgc 900
cgtaccctag gtttgcaggg ttgtgcattc cagtctacta ttgctgagct ccagcgttta 960
aaaatgaagg taggtaaaac ccgggagtct gactacaagg acgacgatga caagtaataa 1020
ggatcc

```

1026

<210> 196

<211> 1389

<212> DNA

<213> Hepatitis E virus

<220>

<223> PLORF32M-14-5.seq

<400> 196

```

gaattcatga ataacatgtc ttttgcacgc cccatgggat caccatgcgc cctagggctg 60
ttctgttgtt gctcttcgtg cttttgccta tgctgcccgc gccacgggcc agccagccgt 120
ctggccgcgc tcgtgggcgg cgtagcggcg gtgccggcgg tggtttctgg ggtgacaggg 180
ttgattctca gcccttcgcc ctcccctata ttcattccaac caacccttc gccgccgatg 240

```

```

tcgtttcaca acccggggct ggaactcgcc ctgcacagcc gccccgcccc cttgggtccg 300
cttggcggtga ccagtcaccag cgcctctccg ctgccccccg tcgtcgatct gccccagctt 360
ggctctgcgcc gcggtgctga tgggactgct gagcttacta ccacagcagc cacacgtttc 420
atgaaggacc tgcacttcgc tggcacgaat ggcgttggtg aggtgggtcg tggatatgcc 480
ctgacactgt tcaatctcgc tgatacgctt ctgcgcggtt taccgacaga attgatttcg 540
tcgggtgggg gccaaactgt ttactccgcg ccggttggtc cagccaatgg cgagccaaca 600
gtaaagttat atacatctgt tgagaatgcg cagcaagaca agggcatcac cattccacat 660
gatatagacc tgggtgactc ccgtgtggtt atccaggatt atgataacca gcatgagcaa 720
gaccgacctc ctccgtcacc tgccccctct cgcctctctc cagttcttcg tgccaatgat 780
gttttgtggc ttccctcac tgccgctgag tatgaccaga ctacgtatgg gtcgtccacc 840
aaccctatgt atgtctctga cacagttacg cttgttaatg tggctactgg tgctcaggct 900
gttgcccgct cccttgattg gtctaaagtt actctggacg gccgccccct tactaccatt 960
cagcagtatt ctaagacatt ttatgttctc ccgtcccgcg ggaagctgtc cttttgggag 1020
gctggcacga ctaaggcccg ctacccttac aattataata ctaccgctag tgaccaaatt 1080
ttgattgaga atgcggcccg ccaccgtgtc gctatttcca cctataccac tagcttaggt 1140
gccggtccta cctcgatctc tgccgctcgc gtactggctc cacactctgc ccttgccggt 1200
cttgaggata ctattgatta ccccgcccggt gcccataact ttgatgattt ttgcccgag 1260
tgccgtaccc taggtttgca gggttgtgca ttccagtcta ctattgctga gctccagcgt 1320
ttaaaaatga aggtaggtaa aacccgggag tctgactaca aggacgacga tgacaagtaa 1380
taaggatcc 1389

```

<210> 197

<211> 74

<212> PRT

<213> Hepatitis E virus

<220>

<223> z12-orf3-5'.pep

<400> 197

```

Met Asn Asn Met Phe Cys Ala Ser Pro Met Gly Ser Pro Cys Ala Leu
  1              5              10              15
Gly Leu Phe Cys Cys Cys Ser Ser Cys Phe Cys Leu Cys Cys Pro Arg
          20              25              30

```

Tyr Asp Asn Gln His Glu Gln Asp Arg Pro Thr Pro Ser Pro Ala Pro
115 120 125

Ser Arg Pro Phe Ser Val Leu Arg Ala Asn Asp Val Leu Trp Leu Ser
 130 135 140
 Leu Thr Ala Ala Glu Tyr Asp Gln Thr Thr Tyr Gly Ser Ser Thr Asn
 145 150 155 160
 Pro Met Tyr Val Ser Asp Thr Val Thr Leu Val Asn Val Ala Thr Gly
 165 170 175
 Ala Gln Ala Val Ala Arg Ser Leu Asp Trp Ser Lys Val Thr Leu Asp
 180 185 190
 Gly Arg Pro Leu Thr Thr Ile Gln Gln Tyr Ser Lys Thr Phe Tyr Val
 195 200 205
 Leu Pro Leu Arg Gly Lys Leu Ser Phe Trp Glu Ala Gly Thr Thr Lys
 210 215 220
 Ala Gly Tyr Pro Tyr Asn Tyr Asn Thr Thr Ala Ser Asp Gln Ile Leu
 225 230 235 240
 Ile Glu Asn Ala Ala Gly His Arg Val Ala Ile Ser Thr Tyr Thr Thr
 245 250 255
 Ser Leu Gly Ala Gly Pro Thr Ser Ile Ser Ala Val Gly Val Leu Ala
 260 265 270
 Pro His Ser Ala Leu Ala Val Leu Glu Asp Thr Ile Asp Tyr Pro Ala
 275 280 285
 Arg Ala His Thr Phe Asp Asp Phe Cys Pro Glu Cys Arg Thr Leu Gly
 290 295 300
 Leu Gln Gly Cys Ala Phe Gln Ser Thr Ile Ala Glu Leu Gln Arg Leu
 305 310 315 320
 Lys Met Lys Val Gly Lys Thr Arg Glu Ser Asp Tyr Lys Asp Asp Asp
 325 330 335

Asp Lys

<210> 200
 <211> 338
 <212> PRT
 <213> Hepatitis E virus

<220>
 <223> plorf2.2-6.pep

<400> 200
 Glu Phe Met Gly Ala Asp Gly Thr Ala Glu Leu Thr Thr Thr Ala Ala
 1 5 10 15
 Thr Arg Phe Met Lys Asp Leu His Phe Ala Gly Thr Asn Gly Val Gly
 20 25 30

Glu Val Gly Arg Gly Ile Ala Leu Thr Leu Phe Asn Leu Ala Asp Thr
 35 40 45
 Leu Leu Gly Gly Leu Pro Thr Glu Leu Ile Ser Ser Ala Gly Gly Gln
 50 55 60
 Leu Phe Tyr Ser Arg Pro Val Val Ser Ala Asn Gly Glu Pro Thr Val
 65 70 75 80
 Lys Leu Tyr Thr Ser Val Glu Asn Ala Gln Gln Asp Lys Gly Ile Thr
 85 90 95
 Ile Pro His Asp Ile Asp Leu Gly Asp Ser Arg Val Val Ile Gln Asp
 100 105 110
 Tyr Asp Asn Gln His Glu Gln Asp Arg Pro Thr Pro Ser Pro Ala Pro
 115 120 125
 Ser Arg Pro Phe Ser Val Leu Arg Ala Asn Asp Val Leu Trp Leu Ser
 130 135 140
 Leu Thr Ala Ala Glu Tyr Asp Gln Thr Thr Tyr Gly Ser Ser Thr Asn
 145 150 155 160
 Pro Met Tyr Val Ser Asp Thr Val Thr Leu Val Asn Val Ala Thr Gly
 165 170 175
 Ala Gln Ala Val Ala Arg Ser Leu Asp Trp Ser Lys Val Thr Leu Asp
 180 185 190
 Gly Arg Pro Leu Thr Thr Ile Gln Gln Tyr Ser Lys Thr Phe Tyr Val
 195 200 205
 Leu Pro Leu Arg Gly Lys Leu Ser Phe Trp Glu Ala Gly Thr Thr Lys
 210 215 220
 Ala Gly Tyr Pro Tyr Asn Tyr Asn Thr Thr Ala Ser Asp Gln Ile Leu
 225 230 235 240
 Ile Glu Asn Ala Ala Gly His Arg Val Ala Ile Ser Thr Tyr Thr Thr
 245 250 255
 Ser Leu Gly Ala Gly Pro Thr Ser Ile Ser Ala Val Gly Val Leu Ala
 260 265 270
 Pro His Ser Ala Leu Ala Val Leu Glu Asp Thr Ile Asp Tyr Pro Ala
 275 280 285
 Arg Ala His Thr Phe Asp Asp Phe Cys Pro Glu Cys Arg Thr Leu Gly
 290 295 300
 Leu Gln Gly Cys Ala Phe Gln Ser Thr Ile Ala Glu Leu Gln Arg Leu
 305 310 315 320
 Lys Met Lys Val Gly Lys Thr Arg Glu Ser Asp Tyr Lys Asp Asp Asp
 325 330 335

Asp Lys

<210> 201

<211> 37

<212> DNA

<213> Hepatitis E virus

<220>

<223> Description of Artificial Sequence: Primer orf35p

<400> 201

tatatgaatt catgaataac atgtcttttg catcgcc 37

<210> 202

<211> 68

<212> DNA

<213> Hepatitis E virus

<220>

<223> Description of Artificial Sequence: Primer orf33p

<400> 202

tatatggatc cttattactt gtcacgctcg tccttgtagt cgcggcgcag accaagctgg 60

ggcagatc 68

<210> 203

<211> 132

<212> PRT

<213> Hepatitis E virus

<220>

<223> pJOorf3-29.pep

<400> 203

Glu Phe Met Asn Asn Met Ser Phe Ala Ser Pro Met Gly Ser Pro Cys
1 5 10 15

Ala Leu Gly Leu Phe Cys Cys Cys Ser Ser Cys Phe Cys Leu Cys Cys
20 25 30

Pro Arg His Arg Pro Ala Ser Arg Leu Ala Ala Val Val Gly Gly Ala
35 40 45

Ala Ala Val Pro Ala Val Val Ser Gly Val Thr Gly Leu Ile Leu Ser
50 55 60

Pro Ser Pro Ser Pro Ile Phe Ile Gln Pro Thr Pro Ser Pro Pro Met
65 70 75 80

Ser Phe His Asn Pro Gly Leu Glu Leu Ala Leu Asp Ser Arg Pro Ala
85 90 95

Pro Leu Ala Pro Leu Gly Val Thr Ser Pro Ser Ala Pro Pro Leu Pro
 100 105 110

Pro Val Val Asp Leu Pro Gln Leu Gly Leu Arg Arg Asp Tyr Lys Asp
 115 120 125

Asp Asp Asp Lys
 130

<210> 204

<211> 132

<212> PRT

<213> Hepatitis E virus

<220>

<223> plorf3-12.pep

<400> 204

Glu Phe Met Asn Asn Met Ser Phe Ala Ser Pro Met Gly Ser Pro Cys
 1 5 10 15

Ala Leu Gly Leu Phe Cys Cys Cys Ser Ser Cys Phe Cys Leu Cys Cys
 20 25 30

Pro Arg His Arg Pro Ala Ser Arg Leu Ala Ala Val Val Gly Gly Ala
 35 40 45

Ala Ala Val Pro Ala Val Val Ser Gly Val Thr Gly Leu Ile Leu Ser
 50 55 60

Pro Ser Pro Ser Pro Ile Phe Ile Gln Pro Thr Pro Ser Pro Pro Met
 65 70 75 80

Ser Phe His Asn Pro Gly Leu Glu Leu Ala Leu Asp Ser Arg Pro Ala
 85 90 95

Pro Leu Ala Pro Leu Gly Val Thr Ser Pro Ser Ala Pro Pro Leu Pro
 100 105 110

Pro Val Val Asp Leu Pro Gln Leu Gly Leu Arg Arg Asp Tyr Lys Asp
 115 120 125

Asp Asp Asp Lys
 130

<210> 205

<211> 48

<212> DNA

<213> Hepatitis E virus

<220>

<223> Description of Artificial Sequence: Primer orf23

<400> 205

48

```
<210> 206
<211> 459
<212> PRT
<213> Hepatitis E virus
```

<220>
<223> CKSORF32M-3.pep

<400> 206															
Glu 1	Phe	Met	Asn	Asn 5	Met	Ser	Phe	Ala	Ser 10	Pro	Met	Gly	Ser	Pro 15	Cys
Ala	Leu	Gly	Leu 20	Phe	Cys	Cys	Cys	Ser 25	Ser	Cys	Phe	Cys	Leu 30	Cys	Cys
Pro	Arg	His 35	Arg	Pro	Ala	Ser	Arg 40	Leu	Ala	Ala	Val	Val 45	Gly	Gly	Val
Ala	Ala 50	Val	Pro	Ala	Val	Val 55	Ser	Gly	Val	Thr	Gly 60	Leu	Ile	Leu	Ser
Pro 65	Ser	Pro	Ser	Pro	Ile 70	Phe	Ile	Gln	Pro	Thr 75	Pro	Ser	Pro	Pro	Met 80
Ser	Phe	His	Asn	Pro 85	Gly	Leu	Glu	Leu	Ala 90	Leu	Asp	Ser	Arg	Pro 95	Ala
Pro	Leu	Ala	Pro 100	Leu	Gly	Val	Thr	Ser 105	Pro	Ser	Ala	Pro	Pro 110	Leu	Pro
Pro	Val 115	Val	Asp	Leu	Pro	Gln	Leu 120	Gly	Leu	Arg	Arg	Gly 125	Ala	Asp	Gly
Thr 130	Ala	Glu	Leu	Thr	Thr	Thr 135	Ala	Ala	Thr	Arg	Phe 140	Met	Lys	Asp	Leu
His 145	Phe	Ala	Gly	Thr 150	Asn	Gly	Val	Gly	Glu 155	Val	Gly	Arg	Gly	Ile 160	Ala
Leu	Thr	Leu	Phe	Asn 165	Leu	Ala	Asp	Thr	Leu 170	Leu	Gly	Gly	Leu	Pro 175	Thr
Glu	Leu	Ile	Ser 180	Ser	Ala	Gly	Gly	Gln 185	Leu	Phe	Tyr	Ser	Arg 190	Pro	Val
Val	Ser	Ala 195	Asn	Gly	Glu	Pro	Thr 200	Val	Lys	Leu	Tyr	Thr 205	Ser	Val	Glu
Asn 210	Ala	Gln	Gln	Asp	Lys	Gly 215	Ile	Thr	Ile	Pro	His 220	Asp	Ile	Asp	Leu
Gly 225	Asp	Ser	Arg	Val 230	Val	Ile	Gln	Asp	Tyr	Asp 235	Asn	Gln	His	Glu	Gln 240

Asp Arg Pro Thr Pro Ser Pro Ala Pro Ser Arg Pro Phe Ser Val Leu
 245 250 255
 Arg Ala Asn Asp Val Leu Trp Leu Ser Leu Thr Ala Ala Glu Tyr Asp
 260 265 270
 Gln Thr Thr Tyr Gly Ser Ser Thr Asn Pro Met Tyr Val Ser Asp Thr
 275 280 285
 Val Thr Leu Val Asn Val Ala Thr Gly Ala Gln Ala Val Ala Arg Ser
 290 295 300
 Leu Asp Trp Ser Lys Val Thr Leu Asp Gly Arg Pro Leu Thr Thr Ile
 305 310 315 320
 Gln Gln Tyr Ser Lys Thr Phe Tyr Val Leu Pro Leu Arg Gly Lys Leu
 325 330 335
 Ser Phe Trp Glu Ala Gly Thr Thr Lys Ala Gly Tyr Pro Tyr Asn Tyr
 340 345 350
 Asn Thr Thr Ala Ser Asp Gln Ile Leu Ile Glu Asn Ala Ala Gly His
 355 360 365
 Arg Val Ala Ile Ser Thr Tyr Thr Thr Ser Leu Gly Ala Gly Pro Thr
 370 375 380
 Ser Ile Ser Ala Val Gly Val Leu Ala Pro His Ser Ala Leu Ala Val
 385 390 395 400
 Leu Glu Asp Thr Ile Asp Tyr Pro Ala Arg Ala His Thr Phe Asp Asp
 405 410 415
 Phe Cys Pro Glu Cys Arg Thr Leu Gly Leu Gln Gly Cys Ala Phe Gln
 420 425 430
 Ser Thr Ile Ala Glu Leu Gln Arg Leu Lys Met Lys Val Gly Lys Thr
 435 440 445
 Arg Glu Ser Asp Tyr Lys Asp Asp Asp Asp Lys
 450 455

<210> 207

<211> 459

<212> PRT

<213> Hepatitis E virus

<220>

<223> PLORF32M-14-5.pep

<400> 207

Glu Phe Met Asn Asn Met Ser Phe Ala Ser Pro Met Gly Ser Pro Cys
 1 5 10 15

Ala Leu Gly Leu Phe Cys Cys Cys Ser Ser Cys Phe Cys Leu Cys Cys
 20 25 30

Pro Arg His Arg Pro Ala Ser Arg Leu Ala Ala Val Val Gly Gly Val
 35 40 45
 Ala Ala Val Pro Ala Val Val Ser Gly Val Thr Gly Leu Ile Leu Ser
 50 55 60
 Pro Ser Pro Ser Pro Ile Phe Ile Gln Pro Thr Pro Ser Pro Pro Met
 65 70 75 80
 Ser Phe His Asn Pro Gly Leu Glu Leu Ala Leu Asp Ser Arg Pro Ala
 85 90 95
 Pro Leu Ala Pro Leu Gly Val Thr Ser Pro Ser Ala Pro Pro Leu Pro
 100 105 110
 Pro Val Val Asp Leu Pro Gln Leu Gly Leu Arg Arg Gly Ala Asp Gly
 115 120 125
 Thr Ala Glu Leu Thr Thr Thr Ala Ala Thr Arg Phe Met Lys Asp Leu
 130 135 140
 His Phe Ala Gly Thr Asn Gly Val Gly Glu Val Gly Arg Gly Ile Ala
 145 150 155 160
 Leu Thr Leu Phe Asn Leu Ala Asp Thr Leu Leu Gly Gly Leu Pro Thr
 165 170 175
 Glu Leu Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro Val
 180 185 190
 Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val Glu
 195 200 205
 Asn Ala Gln Gln Asp Lys Gly Ile Thr Ile Pro His Asp Ile Asp Leu
 210 215 220
 Gly Asp Ser Arg Val Val Ile Gln Asp Tyr Asp Asn Gln His Glu Gln
 225 230 235 240
 Asp Arg Pro Thr Pro Ser Pro Ala Pro Ser Arg Pro Phe Ser Val Leu
 245 250 255
 Arg Ala Asn Asp Val Leu Trp Leu Ser Leu Thr Ala Ala Glu Tyr Asp
 260 265 270
 Gln Thr Thr Tyr Gly Ser Ser Thr Asn Pro Met Tyr Val Ser Asp Thr
 275 280 285
 Val Thr Leu Val Asn Val Ala Thr Gly Ala Gln Ala Val Ala Arg Ser
 290 295 300
 Leu Asp Trp Ser Lys Val Thr Leu Asp Gly Arg Pro Leu Thr Thr Ile
 305 310 315 320
 Gln Gln Tyr Ser Lys Thr Phe Tyr Val Leu Pro Leu Arg Gly Lys Leu
 325 330 335

Ser Phe Trp Glu Ala Gly Thr Thr Lys Ala Gly Tyr Pro Tyr Asn Tyr
 340 345 350

Asn Thr Thr Ala Ser Asp Gln Ile Leu Ile Glu Asn Ala Ala Gly His
 355 360 365

Arg Val Ala Ile Ser Thr Tyr Thr Thr Ser Leu Gly Ala Gly Pro Thr
 370 375 380

Ser Ile Ser Ala Val Gly Val Leu Ala Pro His Ser Ala Leu Ala Val
 385 390 395 400

Leu Glu Asp Thr Ile Asp Tyr Pro Ala Arg Ala His Thr Phe Asp Asp
 405 410 415

Phe Cys Pro Glu Cys Arg Thr Leu Gly Leu Gln Gly Cys Ala Phe Gln
 420 425 430

Ser Thr Ile Ala Glu Leu Gln Arg Leu Lys Met Lys Val Gly Lys Thr
 435 440 445

Arg Glu Ser Asp Tyr Lys Asp Asp Asp Asp Lys
 450 455

<210> 208

<211> 36

<212> DNA

<213> Hepatitis E virus

<220>

<223> Description of Artificial Sequence: Primer
 orf2mid5p

<400> 208

tatatgaatt catgggtgct gatgggactg ctgagc

36

<210> 209

<211> 418

<212> DNA

<213> Hepatitis E virus

<220>

<223> 1440o1.seq

<220>

<221> CDS

<222> (3) .. (416)

<400> 209

ct ggc aty act act gcy att gag cag gct gct ctg gct gcg gcc aat
 Gly Xaa Thr Thr Xaa Ile Glu Gln Ala Ala Leu Ala Ala Ala Asn
 1 5 10 15

47

tcc gcc ttg gcg aat gct gtg gtg gtt cgg ccg ttt tta tcc cgt gtt

95

Ser Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu Ser Arg Val
 20 25 30

caa act gat atc ctt att aac ctg atg caa ccc cgt cag ctt gtg ttc 143
 Gln Thr Asp Ile Leu Ile Asn Leu Met Gln Pro Arg Gln Leu Val Phe
 35 40 45

cgg cct gaa gtt ctc tgg aac cat ccg atc cag cga gtt ata cat aat 191
 Arg Pro Glu Val Leu Trp Asn His Pro Ile Gln Arg Val Ile His Asn
 50 55 60

gag ctg gaa caa tac tgt cga gcc cgc gct ggc cgc tgt ctt gag gtg 239
 Glu Leu Glu Gln Tyr Cys Arg Ala Arg Ala Gly Arg Cys Leu Glu Val
 65 70 75

ggc gct cac cca agg tct att aat gat aac ccc aat gtt ctg cac cgg 287
 Gly Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Leu His Arg
 80 85 90 95

tgc ttt ctc cgc ccg gtt ggg aga gac gtc cag cgc tgg tat tcc gcc 335
 Cys Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp Tyr Ser Ala
 100 105 110

ccc act cgt ggt cca gcg gct aac tgc cgc cgt tct gcg cta cgc ggt 383
 Pro Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu Arg Gly
 115 120 125

ttg ccc cct gtc gac cgc act tac tgt yty gat gg 418
 Leu Pro Pro Val Asp Arg Thr Tyr Cys Xaa Asp
 130 135

<210> 210
 <211> 138
 <212> PRT
 <213> Hepatitis E virus

<400> 210
 Gly Xaa Thr Thr Xaa Ile Glu Gln Ala Ala Leu Ala Ala Ala Asn Ser
 1 5 10 15

Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu Ser Arg Val Gln
 20 25 30

Thr Asp Ile Leu Ile Asn Leu Met Gln Pro Arg Gln Leu Val Phe Arg
 35 40 45

Pro Glu Val Leu Trp Asn His Pro Ile Gln Arg Val Ile His Asn Glu
 50 55 60

Leu Glu Gln Tyr Cys Arg Ala Arg Ala Gly Arg Cys Leu Glu Val Gly
 65 70 75 80

Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Leu His Arg Cys
 85 90 95

Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp Tyr Ser Ala Pro

100

105

110

Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu Arg Gly Leu
 115 120 125

Pro Pro Val Asp Arg Thr Tyr Cys Xaa Asp
 130 135

<210> 211

<211> 197

<212> DNA

<213> Hepatitis E virus

<220>

<223> 1440o2.seq

<220>

<221> CDS

<222> (2)..(196)

<400> 211

g aca gaa ttr att tcg tcg gct gga ggt caa ctg ttc tac tcc cgc ccg 49
 Thr Glu Xaa Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro
 1 5 10 15

ggt gtc tca gcc aat ggc gag ccg act gtt aag tta tac acc tct gtc 97
 Val Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val
 20 25 30

gag aat gca cag cag gat aag ggc att gct ata cca cat gat ata gac 145
 Glu Asn Ala Gln Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp
 35 40 45

tta ggg gat tcc cgt gtg gtt ata caa gat tat gay aac car cay gaa 193
 Leu Gly Asp Ser Arg Val Val Ile Gln Asp Tyr Xaa Asn Xaa Xaa Glu
 50 55 60

caa g 197
 Gln
 65

<210> 212

<211> 65

<212> PRT

<213> Hepatitis E virus

<400> 212

Thr Glu Xaa Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro
 1 5 10 15

Val Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val
 20 25 30

Glu Asn Ala Gln Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp
 35 40 45

Leu Gly Asp Ser Arg Val Val Ile Gln Asp Tyr Xaa Asn Xaa Xaa Glu
 50 55 60

Gln
 65

<210> 213
 <211> 418
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> 2015-1.seq

<220>
 <221> CDS
 <222> (3) .. (416)

<400> 213
 ct ggc aty act act gcy att gag cag gct gct ctg gct gcg gct aac 47
 Gly Xaa Thr Thr Xaa Ile Glu Gln Ala Ala Leu Ala Ala Ala Asn
 1 5 10 15
 tct gcc ttg gcg aat gct gtg gtg gtc cgg ccg ttc ctg tcc cgc act 95
 Ser Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu Ser Arg Thr
 20 25 30
 cag act gat att ctt att aat ttg atg caa ccc cgg caa ctt gta ttc 143
 Gln Thr Asp Ile Leu Ile Asn Leu Met Gln Pro Arg Gln Leu Val Phe
 35 40 45
 cgc cct gag gtt ttg tgg aac cat ccg atc cag cga gtc ata cat aat 191
 Arg Pro Glu Val Leu Trp Asn His Pro Ile Gln Arg Val Ile His Asn
 50 55 60
 gag ctg gag cag tat tgc cgt gct cgt gct ggt cgc tgc ctg gag gtt 239
 Glu Leu Glu Gln Tyr Cys Arg Ala Arg Ala Gly Arg Cys Leu Glu Val
 65 70 75
 ggg gct cat cca aga tct atc aat gac aac cct aat gtt ctg cac cgg 287
 Gly Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Leu His Arg
 80 85 90 95
 tgt ttc ctg cgt ccg gtt ggg cga gac gta cag cgt tgg tat tct gcc 335
 Cys Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp Tyr Ser Ala
 100 105 110
 cct act cgc ggc ccg gcg gct aat tgc cgc cgt tcc gcg tta cgt ggc 383
 Pro Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu Arg Gly
 115 120 125
 cta cct cct gtc gac cgc act tac tgt yty gat gg 418
 Leu Pro Pro Val Asp Arg Thr Tyr Cys Xaa Asp
 130 135

<210> 214
 <211> 138
 <212> PRT
 <213> Hepatitis E virus

<400> 214
 Gly Xaa Thr Thr Xaa Ile Glu Gln Ala Ala Leu Ala Ala Ala Asn Ser
 1 5 10 15
 Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu Ser Arg Thr Gln
 20 25 30
 Thr Asp Ile Leu Ile Asn Leu Met Gln Pro Arg Gln Leu Val Phe Arg
 35 40 45
 Pro Glu Val Leu Trp Asn His Pro Ile Gln Arg Val Ile His Asn Glu
 50 55 60
 Leu Glu Gln Tyr Cys Arg Ala Arg Ala Gly Arg Cys Leu Glu Val Gly
 65 70 75 80
 Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Leu His Arg Cys
 85 90 95
 Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp Tyr Ser Ala Pro
 100 105 110
 Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu Arg Gly Leu
 115 120 125
 Pro Pro Val Asp Arg Thr Tyr Cys Xaa Asp
 130 135

<210> 215
 <211> 197
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> 2015o2.seq

<220>
 <221> CDS
 <222> (2)..(196)

<400> 215
 g aca gaa ttr att tcg tcg gct gga ggc cag ctc ttc tac tcc cgc cca 49
 Thr Glu Xaa Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro
 1 5 10 15
 gtc gtc tca gcc aat ggc gag ccg act gtt aaa ttg tat aca tcc gtc 97
 Val Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val
 20 25 30
 gag aat gcg cag cag gac aag ggc att gcc ata cca cat gat ata gat 145

Glu Asn Ala Gln Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp
 35 40 45

cta gga gat tcc cgc gtg gtt atc cag gat tat gay aac car cay gaa 193
 Leu Gly Asp Ser Arg Val Val Ile Gln Asp Tyr Xaa Asn Xaa Xaa Glu
 50 55 60

caa g 197
 Gln
 65

<210> 216
 <211> 65
 <212> PRT
 <213> Hepatitis E virus

<400> 216
 Thr Glu Xaa Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro
 1 5 10 15

Val Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val
 20 25 30

Glu Asn Ala Gln Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp
 35 40 45

Leu Gly Asp Ser Arg Val Val Ile Gln Asp Tyr Xaa Asn Xaa Xaa Glu
 50 55 60

Gln
 65

<210> 217
 <211> 251
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> 14404-2.seq

<220>
 <221> CDS
 <222> (3)..(251)
 <223> orf2

<220>
 <223> orf3 from position 1 to position 165

<400> 217
 at att cat cca acc aac ccc ttt gcc tcc gac gtc gta tcg caa tcc 47
 Ile His Pro Thr Asn Pro Phe Ala Ser Asp Val Val Ser Gln Ser
 1 5 10 15

ggg gct gga gct cgc cct cga cag ccg gcc cgc ccc ctc ggc tcc tct 95
 Gly Ala Gly Ala Arg Pro Arg Gln Pro Ala Arg Pro Leu Gly Ser Ser

	20	25	30	
tgg cgt gac cag tcc cag cgc ccc ccc gct gtc ccc cgt cgt cga tct				143
Trp Arg Asp Gln Ser Gln Arg Pro Pro Ala Val Pro Arg Arg Arg Ser				
	35	40	45	
acc cca act ggg gct gcg ccg cta act gct gtt tca cca gcg cct gat				191
Thr Pro Thr Gly Ala Ala Pro Leu Thr Ala Val Ser Pro Ala Pro Asp				
	50	55	60	
acg gcc cca gtc cct gat gtt gac tct cgt ggc gct atc ttg cgc cgg				239
Thr Ala Pro Val Pro Asp Val Asp Ser Arg Gly Ala Ile Leu Arg Arg				
	65	70	75	
cag tat aac cta				251
Gln Tyr Asn Leu				
80				

<210> 218
 <211> 83
 <212> PRT
 <213> Hepatitis E virus

<400> 218
 Ile His Pro Thr Asn Pro Phe Ala Ser Asp Val Val Ser Gln Ser Gly
 1 5 10 15
 Ala Gly Ala Arg Pro Arg Gln Pro Ala Arg Pro Leu Gly Ser Ser Trp
 20 25 30
 Arg Asp Gln Ser Gln Arg Pro Pro Ala Val Pro Arg Arg Arg Ser Thr
 35 40 45
 Pro Thr Gly Ala Ala Pro Leu Thr Ala Val Ser Pro Ala Pro Asp Thr
 50 55 60
 Ala Pro Val Pro Asp Val Asp Ser Arg Gly Ala Ile Leu Arg Arg Gln
 65 70 75 80
 Tyr Asn Leu

<210> 219
 <211> 55
 <212> PRT
 <213> Hepatitis E virus

<220>
 <223> 14404-2.seq orf3

<400> 219
 Ile Phe Ile Gln Pro Thr Pro Leu Pro Pro Thr Ser Tyr Arg Asn Pro
 1 5 10 15
 Gly Leu Glu Leu Ala Leu Asp Ser Arg Pro Ala Pro Ser Ala Pro Leu

20 25 30
 Gly Val Thr Ser Pro Ser Ala Pro Pro Leu Ser Pro Val Val Asp Leu
 35 40 45

Pro Gln Leu Gly Leu Arg Arg
 50 55

<210> 220
 <211> 251
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> 20154-2.seq

<220>
 <221> CDS
 <222> (3) .. (251)
 <223> orf2

<220>
 <223> orf3 from position 1 to position 165

<400> 220
 at att cat cca acc aac ccc ttt gcc gcc gac gtc gta tca caa ccc 47
 Ile His Pro Thr Asn Pro Phe Ala Ala Asp Val Val Ser Gln Pro
 1 5 10 15

 ggg gct gga gct cgc cct cga cag ccg ccc cgc ccc ctc ggc tcc tct 95
 Gly Ala Gly Ala Arg Pro Arg Gln Pro Pro Arg Pro Leu Gly Ser Ser
 20 25 30

 tgg cgt gat cag tcc cag cgc ccc tcc gct gcc ccc cgt cgt cga tct 143
 Trp Arg Asp Gln Ser Gln Arg Pro Ser Ala Ala Pro Arg Arg Arg Ser
 35 40 45

 acc cca gct ggg gct gcg ccg tta act gct gtt tcc cct gcg ccc gat 191
 Thr Pro Ala Gly Ala Ala Pro Leu Thr Ala Val Ser Pro Ala Pro Asp
 50 55 60

 acg gcc cca gtc ccc gac gtt gat tcc cgt ggt gcc atc ctg cgc cgg 239
 Thr Ala Pro Val Pro Asp Val Asp Ser Arg Gly Ala Ile Leu Arg Arg
 65 70 75

 cag tat aac cta 251
 Gln Tyr Asn Leu
 80

<210> 221
 <211> 83
 <212> PRT
 <213> Hepatitis E virus

<400> 221

Ile His Pro Thr Asn Pro Phe Ala Ala Asp Val Val Ser Gln Pro Gly
 1 5 10 15
 Ala Gly Ala Arg Pro Arg Gln Pro Pro Arg Pro Leu Gly Ser Ser Trp
 20 25 30
 Arg Asp Gln Ser Gln Arg Pro Ser Ala Ala Pro Arg Arg Arg Ser Thr
 35 40 45
 Pro Ala Gly Ala Ala Pro Leu Thr Ala Val Ser Pro Ala Pro Asp Thr
 50 55 60
 Ala Pro Val Pro Asp Val Asp Ser Arg Gly Ala Ile Leu Arg Arg Gln
 65 70 75 80
 Tyr Asn Leu

<210> 222
 <211> 55
 <212> PRT
 <213> Hepatitis E virus

<220>
 <223> 20154-2.seq orf3

<400> 222
 Ile Phe Ile Gln Pro Thr Pro Leu Pro Pro Thr Ser Tyr His Asn Pro
 1 5 10 15
 Gly Leu Glu Leu Ala Leu Asp Ser Arg Pro Ala Pro Ser Ala Pro Leu
 20 25 30
 Gly Val Ile Ser Pro Ser Ala Pro Pro Leu Pro Pro Val Val Asp Leu
 35 40 45
 Pro Gln Leu Gly Leu Arg Arg
 50 55

<210> 223
 <211> 48
 <212> PRT
 <213> Hepatitis E virus

<220>
 <223> US-2 3-2e

<400> 223
 Thr Ile Asp Tyr Pro Ala Arg Ala His Thr Phe Asp Asp Phe Cys Pro
 1 5 10 15
 Glu Cys Arg Thr Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr Ile
 20 25 30
 Ala Glu Leu Gln Arg Leu Lys Met Lys Val Gly Lys Thr Arg Glu Ser

35

40

45

<210> 224

<211> 33

<212> PRT

<213> Hepatitis E virus

<220>

<223> US-2 4-2

<400> 224

Asp	Ser	Arg	Pro	Ala	Pro	Leu	Val	Pro	Leu	Gly	Val	Thr	Ser	Pro	Ser
1				5					10					15	

Ala	Pro	Pro	Leu	Pro	Pro	Val	Val	Asp	Leu	Pro	Gln	Leu	Gly	Leu	Arg
			20					25					30		

Arg

<210> 225

<211> 450

<212> DNA

<213> Hepatitis E virus

<220>

<223> 5p.pile {hpesvp}

<400> 225

```

ggctcctggc atcactactg ctattgagca ggctgctcta gcagcggcca actctgccct 60
ggcgaatgct gtggtagtta ggccctttct ctctcaccag cagattgaga tcctcattaa 120
cctaatagcaa cctcgccagc ttgttttccg ccccgagggt ttctggaatc atcccatcca 180
gcgtgtcatc cataacgagc tggagcttta ctgcgcgcgc cgctccggcc gctgtcttga 240
aattggcgcc catccccgct caataaatga taatcctaata gtggtcacc gctgcttcct 300
ccgccttggt gggcgtgatg ttcagcgctg gtataactgct cccactcgcg ggccggctgc 360
taattgccgg cgttcgcgcg tgcgcggggt tcccgctgct gaccgcactt actgcctcga 420
cgggttttct ggctgtaact ttcccgcoga

```

450

<210> 226

<211> 450

<212> DNA

<213> Hepatitis E virus

<220>

<223> 5p.pile {hpeuigh}

<400> 226

```

ggctcctggc atcactactg ctattgagca ggctgctcta gcagcggcca attctgcct 60
tgcgaatgct gtggtagtta ggccttttct ctctcaccag cagattgaga tccttattaa 120
cctaatagcaa cctcgccagc ttgttttccg ccccgagggt ttctggaacc accccatcca 180
gcgtgtcatc cataatgagc tggagcttta ctgtcgcgcc cgctccggcc gctgccttga 240
aattgggtgcc caccctcgt caataaacga caatccta atgtgggtccacc gctgcttcct 300
ccgccctgcc gggcgtgatg ttcagcggtg gtatactgct cctaccgcgc gcccggtgc 360
taattgccgc ggttccgcac tgcgcgggct ccccgctgct gaccgcactt actgcttcga 420
cgggttttct ggctgtaact ttccgcgcga 450

```

<210> 227

<211> 450

<212> DNA

<213> Hepatitis E virus

<220>

<223> 5p.pile {hpea}

<400> 227

```

ggctcctggc atcactactg ctattgagca ggctgctcta gcagcggcca actctgcct 60
tgcgaatgct gtggtagtta ggccttttct ctctcaccag cagattgaga tccttattaa 120
cctaatagcaa cctcgccagc ttgttttccg ccccgagggt ttctggaacc atcccatcca 180
gcgtgttatc cataatgagc tggagcttta ctgtcgcgcc cgctccggcc gctgcctega 240
aattgggtgcc ccccccgct caataaatga caatccta atgtgggtccacc gttgcttcct 300
ccgtcctgcc gggcgtgatg ttcagcggtg gtatactgcc cctaccgcgc gcccggtgc 360
taattgccgc cgttccgcgc tgcgcgggct ccccgctgct gaccgcactt actgcttcga 420
cgggttttct ggctgtaact ttccgcgcga 450

```

<210> 228

<211> 446

<212> DNA

<213> Hepatitis E virus

<220>

<223> 5p.pile {840455p}

<400> 228

```

cctggcatta ctactgcat tgagcaggct gctctggctg eggccaattc tgccttggcg 60

```

```

aatgctgtgg tggttcggcc gtttttatct cgcgtgcaaa ccgagattct tattaatttg 120
atgcaacccc ggcagttggt tttccgccct gaggtacttt ggaatcaccc tatccagcgg 180
gttatacata atgaattaga acagtactgc cgggctcggg ctggtcgttg cttggagggt 240
ggagctcacc caagatccat taatgacaac cccaacgttc tgcacgggtg tttccttaga 300
ccggttggcc gagatgttca gcgctggtac tctgccccca cccgcggccc tgcgggcta 360
tgccgcgcgt ccgcgttgcg tgggtctccc cccgctgacc gcacttactg ctttgatgga 420
ttctcccggt gtgcttttgc tgcaga 446

```

```

<210> 229
<211> 450
<212> DNA
<213> Hepatitis E virus

```

```

<220>
<223> 5p.pile {hpenssp}

```

```

<400> 229
ggctcctggc atcactactg ctattgagca agcagctcta gcagcggcca actccgccct 60
tgcaaatgct gtggtggtcc ggcctttcct tcccacag caggttgaga tccttataaa 120
tctcatgcaa cctcggcagc tgggtgttct tctgaggtt ttttgaatc acccgattca 180
acgtgttata cataatgagc ttgagcagta ttgccgtgct cgctcgggtc gctgccttga 240
gattggagcc caccacgcgt ccattaatga taatccta gtcctccatc gctgctttct 300
ccaccccgtc ggccgggatg ttcagcgtg gtacacagcc ccgactaggg gacctgcggc 360
gaactgtcgc cgctcggcac ttcggtgtct gccaccagcc gaccgcactt actgttttga 420
tggttttgcc ggctgccgtt ttgccgcga 450

```

```

<210> 230
<211> 450
<212> DNA
<213> Hepatitis E virus

```

```

<220>
<223> 5p Consensus

```

```

<220>
<221> variation
<222> ()..(450)
<223> The nucleotide identity of each n is indicated in
      Figure 9.

```

```

<400> 230
nnnnctggc atnactactg cnattgagca ngcngctctn gcngcggcca antcngccnt 60

```

ngcgaatgct gtggtngttn ggcntttnt ntncnnnng cannnngaga tncnatnaa 120
 nntnatgcaa ccncgncagn tngtnttneg necngaggtn ntntggaanc anccnatnca 180
 ncnngtnatn cataangann tngancnnta ntgncngcn cgnnncggnc gntgnntnga 240
 nnttgngcn canccnngnt cnatnaanga naancnaaan gtntncanc gntgnttnt 300
 nnnncngnn ggncnggatg ttcagcgtg gtannngcn ccnacnngng gncngcngc 360
 naantgncgn ngntcngcnn tncngggnet necnnngcn gaccgcactt actgnntnga 420
 nggnttnncn ngntgnnnt ttnengnga 450

<210> 231

<211> 300

<212> DNA

<213> Hepatitis E virus

<220>

<223> 3p.pile {hpea} shown in Figure 9B

<400> 231

actgagtcag tgaagccagt gcttgacctg acaaattcaa ttctgtgtcg ggtggaatga 60
 ataacatgtc ttttgctgcg cccatgggtt cgcgaccatg cgccctcggc ctattttgct 120
 gttgctctc atgtttctgc ctatgtgtcc cgcgccaccg cccggtcagc cgtctggccg 180
 ccgtcgtggg cggcgcagcg gcggttcggg cgggtggttc tggggtgacc gggttgattc 240
 tcagcccttc gcaatcccct atattcatcc aaccaacccc ttgcgcccg atgtcacgc 300

<210> 232

<211> 300

<212> DNA

<213> Hepatitis E virus

<220>

<223> 3p.pile {hpeuigh} shown in Figure 9B

<400> 232

actgagtcgg tgaagccagt gctcgacttg acaaattcaa tctgtgtcg ggtggaatga 60
 ataacatgtc ttttgctgcg cccatgggtt ggcgaccatg cgccctcggc ctattttgct 120
 gttgctctc atgtttctgc ctatgtgtcc cgcgccaccg cccggtcagc cgtctggccg 180
 ccgtcgtggg cggcgcagcg gcggttcggg cgggtggttc tggggtgacc gggttgattc 240
 tcagcccttc gcaatcccct atattcatcc aaccaacccc ttgcgcccg atgtcacgc 300

<210> 233

<211> 300
<212> DNA
<213> Hepatitis E virus

<220>
<223> 3p.pile {hpesvp} shown in Figure 9B

<400> 233
actgagtcag taaaaccagt gctcgacttg acaaattcaa tcttggtgctg ggtggaatga 60
ataacatgtc ttttgctgcg cccatgggtt cgcgaccatg cgccctcggc ctattttggtt 120
gctgctcctc atgtttttgc ctatgctgcc cgcgccaccg cccggtcagc cgtctggccg 180
ccgtcgtggg cggcgcagcg gcgggtccgg cgggtggtttc tggggtgacc gggttgattc 240
tcagcccttc gcaatcccct atattcatcc aaccaacccc ttcgcccccg atgtcaccgc 300

<210> 234
<211> 300
<212> DNA
<213> Hepatitis E virus

<220>
<223> 3p.pile {hpenssp} shown in Figure 9B

<400> 234
acagagtctg ttaagcctat acttgacctt acacactcaa ttatgcaccg gtctgaatga 60
ataacatgtg gtttgctgcg cccatgggtt cgccaccatg cgccctaggc ctcttttgct 120
gttgctcctc ttgtttctgc ctatgttgcc cgcgccaccg accggtcagc cgtctggccg 180
ccgtcgtggg cggcgcagcg gcgggtaccg cgggtggtttc tggggtgacc gggttgattc 240
tcagcccttc gcaatcccct atattcatcc aaccaacccc tttgccccag acgttgccgc 300

<210> 235
<211> 297
<212> DNA
<213> Hepatitis E virus

<220>
<223> 3p.pile {840453p} shown in Figure 9B

<400> 235
acagagacta ttaaacctgt acttgatctc acaaattcca tcatacagcg ggtggaatga 60
ataacatgtc ttttgcatcg cccatgggat caccatgcgc cctagggctg ttctgttggtt 120
gttcctcatg tttctgccta tgctgcccgc gccaccggcc ggtcagccgt ctggccgctg 180
ccgtggggcg cgacgcggcg gtgcggcgcg tggtttctgg agtgacaggg ttgattctca 240
gcccttcgcc ctcccctata ttcattccaac caaccccttc gccgccgatg tcgtttc 297

<210> 236
 <211> 300
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> 3p Consensus shown in Figure 9B

<220>
 <221> variation
 <222> (1)..(300)
 <223> The nucleotide identity of each n is indicated in
 Figure 9B

<400> 236
 acngagncnn tnaanccnnt nctnganntn acanantcna tnntnnnnnecg gnnngaataga 60
 ataacatgtn ntttgcnnecg cccatgggnt nnnnaccatg cgccttnggn ctntttntgnt 120
 gntgntcctc ntgtttntgc ctatnntgcc cgcgccaccg nccggtcagc cgtctggccg 180
 ncgncgtggg cggcgcagcg gcggtnccgg cggtggtttc tggngtgacn gggttgattc 240
 tcagcccttc gcnnccccct atattcatcc aaccaacccc ttngccncng angtnnnnnnc 300

<210> 237
 <211> 250
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> 3p.pile {hpea} shown in Figure 9C

<400> 237
 agcgcttacc ctgtttaacc ttgctgacac cctgcttggc ggtctaccga cagaattgat 60
 ttcgtcggct ggtggccagc tgttctactc tcgccccgtc gtctcagcca atggcgagcc 120
 gactgttaag ctgtatacat ctgtggagaa tgctcagcag gataagggtg ttgcaatccc 180
 gcatgacatc gacctcgggg aatcccgtgt agttattcag gattatgaca accaacaatga 240
 gcaggaccga 250

<210> 238
 <211> 250
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> 3p.pile {hpeuigh} shown in Figure 9C

<400> 238

```

agcgcttacc ctgtttaacc ttgctgacac cctgcttggc ggtctaccga cagaattgat 60
ttcgtcggct ggtggccagc tgttctactc tcgccccgtc gtctcagcca atggcgagcc 120
gactgttaag ctgtatacat ctgtagagaa tgctcagcag gataagggtg ttgcaatccc 180
gcatgacatc gacctcgggg aatctcgagt tgttattcag gattatgaca accaacaatga 240
gcaggaccgg                                     250

```

```

<210> 239
<211> 250
<212> DNA
<213> Hepatitis E virus

```

```

<220>
<223> 3p.pile {hpesvp} shown in Figure 9C

```

```

<400> 239
agccctcacc ctgttcaacc ttgctgacac tctgcttggc ggcttgccga cagaattgat 60
ttcgtcggct ggtggccagc tgttctactc ccgtcccgtt gtctcagcca atggcgagcc 120
gactgttaag ttgtatacat ctgtagagaa tgctcagcag gataagggtg ttgcaatccc 180
gcatgacatt gacctcggag aatctcgtgt gggtattcag gattatgata accaacaatga 240
acaagatcgg                                     250

```

```

<210> 240
<211> 250
<212> DNA
<213> Hepatitis E virus

```

```

<220>
<223> 3p.pile {hpenssp} shown in Figure 9C

```

```

<400> 240
agctctaaca ttacttaacc ttgctgacac gctcctcggc gggctcccga cagaattaat 60
ttcgtcggct ggcgggcaac tgttttattc ccgcccgtt gtctcagcca atggcgagcc 120
aaccgtgaag ctctatacat cagtggagaa tgctcagcag gataagggtg ttgctatccc 180
ccacgatatc gatcttggtg attcgcgtgt ggtcattcag gattatgaca accagcatga 240
gcaggatcgg                                     250

```

```

<210> 241
<211> 250
<212> DNA
<213> Hepatitis E virus

```

```

<220>

```

<223> 3p.pile {840453p} shown in Figure 9C

<400> 241

```

tgccctgact ctgtttaatc ttgctgatac gcttcttggt ggtttaccga cagaattgat 60
ttcgtcggct ggggggtcaac tgttttactc ccgccctggt cagaattgat ttcgtcggct 120
gggggtcaac tgttttactc ccgccctggt tgcgcagcaa gacaaggga tcaccattcc 180
acacgacata gatttaggtg actocctgtg gggtatccag gattatgata accagcacga 240
acaagatcga                                     250

```

<210> 242

<211> 250

<212> DNA

<213> Hepatitis E virus

<220>

<223> 3p Consensus shown in Figure 9C

<220>

<221> variation

<222> ()..(250)

<223> The nucleotide identity of each n is indicated in
Figure 9C

<400> 242

```

ngcnctnacn ntnntnaanc ttgctganac nctnctnggn ggnntnccga cagaattnat 60
ttcgtcggct ggnggncanc tgtntantc ncgncngtn gtctengcca atggcgagcc 120
nacngtnaag ntntanacat cngtngagaa tgencagcan ganaagggnn tnnchnatncc 180
ncanganatn ganntnggng antcncnggt ngtnatncag gattatgana accancanga 240
ncanganecn                                     250

```

<210> 243

<211> 418

<212> DNA

<213> Hepatitis E virus

<220>

<223> Aulol-wlabolpl.pat

<220>

<221> CDS

<222> (3)..(416)

<400> 243

```

ct ggc aty act act gcy att gag caa gct gct ctg gct gcg gcc aat 47
  Gly Xaa Thr Thr Xaa Ile Glu Gln Ala Ala Leu Ala Ala Ala Asn
    1             5             10             15

```

tct gcc ttg gcg aat gct gtg gtg gtt cgg ccg ttt tta tcc cgt gtg	95
Ser Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu Ser Arg Val	
20 25 30	
cag act gag atc ctt att aac ttg atg caa cct cgg cag ctg gtg ttc	143
Gln Thr Glu Ile Leu Ile Asn Leu Met Gln Pro Arg Gln Leu Val Phe	
35 40 45	
cga cct gag gtg ctt tgg aat cat ccc att cag cgg gtt atc cat aat	191
Arg Pro Glu Val Leu Trp Asn His Pro Ile Gln Arg Val Ile His Asn	
50 55 60	
gag tta gaa caa tac tgc cgg gcc cgg gcc ggc cgt tgc cta gag gtg	239
Glu Leu Glu Gln Tyr Cys Arg Ala Arg Ala Gly Arg Cys Leu Glu Val	
65 70 75	
ggg gcc cac cca agg tcc att aac gat aac ccc aat gtt ttg cac cgg	287
Gly Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Leu His Arg	
80 85 90 95	
tgt ttt ctg cga ccg gtc ggg agg gat gtt cag cgc tgg tac tct gcc	335
Cys Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp Tyr Ser Ala	
100 105 110	
ccc acc cgc ggc cct gcg gct aac tgc cgc cgc tcc gct ttg cgt ggc	383
Pro Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu Arg Gly	
115 120 125	
ctt ccc ccc gtc gac cgc act tac tgt yty gat gg	418
Leu Pro Pro Val Asp Arg Thr Tyr Cys Xaa Asp	
130 135	
<210> 244	
<211> 138	
<212> PRT	
<213> Hepatitis E virus	
<400> 244	
Gly Xaa Thr Thr Xaa Ile Glu Gln Ala Ala Leu Ala Ala Ala Asn Ser	
1 5 10 15	
Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu Ser Arg Val Gln	
20 25 30	
Thr Glu Ile Leu Ile Asn Leu Met Gln Pro Arg Gln Leu Val Phe Arg	
35 40 45	
Pro Glu Val Leu Trp Asn His Pro Ile Gln Arg Val Ile His Asn Glu	
50 55 60	
Leu Glu Gln Tyr Cys Arg Ala Arg Ala Gly Arg Cys Leu Glu Val Gly	
65 70 75 80	
Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Leu His Arg Cys	
85 90 95	

Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp Tyr Ser Ala Pro
 100 105 110

Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu Arg Gly Leu
 115 120 125

Pro Pro Val Asp Arg Thr Tyr Cys Xaa Asp
 130 135

<210> 245

<211> 197

<212> DNA

<213> Hepatitis E virus

<220>

<223> Aulo2-wlao2.pat

<220>

<221> CDS

<222> (2) .. (196)

<400> 245

g aca gaa ttr att tcg tcg gct ggg gga cag tta ttc tac tcc cgc cct 49
 Thr Glu Xaa Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro
 1 5 10 15

gty gtc tca gcc aat ggc gag ccg act gtt aaa tta tat aca tct gta 97
 Xaa Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val
 20 25 30

gag aat gcg cag cag gac aag ggg att gcc atc cca cat gat ata gat 145
 Glu Asn Ala Gln Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp
 35 40 45

ctg ggc gac tct cgt gtg gtg atc cag gat tat gay aac car cay gaa 193
 Leu Gly Asp Ser Arg Val Val Ile Gln Asp Tyr Xaa Asn Xaa Xaa Glu
 50 55 60

caa g 197
 Gln
 65

<210> 246

<211> 65

<212> PRT

<213> Hepatitis E virus

<400> 246

Thr Glu Xaa Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro
 1 5 10 15

Xaa Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val
 20 25 30

Glu Asn Ala Gln Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp

35

40

45

Leu Gly Asp Ser Arg Val Val Ile Gln Asp Tyr Xaa Asn Xaa Xaa Glu
 50 55 60

Gln
 65

<210> 247

<211> 418

<212> DNA

<213> Hepatitis E virus

<220>

<223> Arl01- f73olpl.pat

<220>

<221> CDS

<222> (3) .. (416)

<400> 247

ct ggc aty act act gcy att gag caa gct gct ctg gct gcg gcc aac 47
 Gly Xaa Thr Thr Xaa Ile Glu Gln Ala Ala Leu Ala Ala Ala Asn
 1 5 10 15

tct gcc ttg gcg aat gct gtg gtg gtt cgg ccg ttt tta tcc cgt gtg 95
 Ser Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu Ser Arg Val
 20 25 30

cag acc gag att ctt att aac cta atg caa ccc cgg cag ctg gtt ttt 143
 Gln Thr Glu Ile Leu Ile Asn Leu Met Gln Pro Arg Gln Leu Val Phe
 35 40 45

cgt cct gag gtg ctt tgg aac cat cct atc cag cgg gtt att cat aat 191
 Arg Pro Glu Val Leu Trp Asn His Pro Ile Gln Arg Val Ile His Asn
 50 55 60

gag tta gaa cag tac tgt cgg gct cgg gct ggt cgc tgc cta gag gtc 239
 Glu Leu Glu Gln Tyr Cys Arg Ala Arg Ala Gly Arg Cys Leu Glu Val
 65 70 75

ggg gcc cac cca agg tcc att aat gat aac cct aat gtt ttg cac cgg 287
 Gly Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Leu His Arg
 80 85 90 95

tgc ttc cta cga cca gtc ggg agg gat gtt caa cgt tgg tat tcc gcc 335
 Cys Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp Tyr Ser Ala
 100 105 110

ccc acc cgc ggt cct gct gcc aac tgc cgc cgt tcc gct ctg cgc ggc 383
 Pro Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu Arg Gly
 115 120 125

ctc cct ccc gtc gac cgc act tac tgt yty gat gg 418
 Leu Pro Pro Val Asp Arg Thr Tyr Cys Xaa Asp
 130 135

<210> 248
 <211> 138
 <212> PRT
 <213> Hepatitis E virus

<400> 248
 Gly Xaa Thr Thr Xaa Ile Glu Gln Ala Ala Leu Ala Ala Ala Asn Ser
 1 5 10 15
 Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu Ser Arg Val Gln
 20 25 30
 Thr Glu Ile Leu Ile Asn Leu Met Gln Pro Arg Gln Leu Val Phe Arg
 35 40 45
 Pro Glu Val Leu Trp Asn His Pro Ile Gln Arg Val Ile His Asn Glu
 50 55 60
 Leu Glu Gln Tyr Cys Arg Ala Arg Ala Gly Arg Cys Leu Glu Val Gly
 65 70 75 80
 Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Leu His Arg Cys
 85 90 95
 Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp Tyr Ser Ala Pro
 100 105 110
 Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu Arg Gly Leu
 115 120 125
 Pro Pro Val Asp Arg Thr Tyr Cys Xaa Asp
 130 135

<210> 249
 <211> 145
 <212> DNA
 <213> Hepatitis E virus

<220>
 <223> Arl-f73o2p2.pat

<220>
 <221> CDS
 <222> (1) .. (144)

<400> 249
 gty gtc tcr gcc aat ggc gag ccg act gtt aag cta tat aca tct gta 48
 Xaa Val Xaa Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val
 1 5 10 15
 gag aac gcg cag cag gat aaa ggg atc gcc att cca cac gat ata gat 96
 Glu Asn Ala Gln Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp
 20 25 30

<400>	251
ct ggc aty act act gcy att gag caa gct gct ctg gct gcg gct aac Gly Xaa Thr Thr Xaa Ile Glu Gln Ala Ala Leu Ala Ala Asn	47
1 5 10 15	
tct gcc ttg gcg aat gct gtg gtg gtt cgg ccg ttt cta tcc cgt gtg Ser Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu Ser Arg Val	95
20 25 30	
cag act gag atc ctt att aac tta atg car ccc cgg car ctg gtt ttc Gln Thr Glu Ile Leu Ile Asn Leu Met Xaa Pro Arg Xaa Leu Val Phe	143
35 40 45	
cgt ccc gag gtg ctt tgg aat cat ccc att caa cgg gtt att cat aat Arg Pro Glu Val Leu Trp Asn His Pro Ile Gln Arg Val Ile His Asn	191
50 55 60	
gaa tta gag cag tac tgc cgg acc cgg gct ggc cgt tgt tta gag gtc Glu Leu Glu Gln Tyr Cys Arg Thr Arg Ala Gly Arg Cys Leu Glu Val	239
65 70 75	
gga gcc cat cca agg tcc att aat gac aac cct aac gtt cyg cac cgg Gly Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Xaa His Arg	287
80 85 90 95	

tgc ttc tta cga cca gtc ggg agg gat gtc caa cga tgg tac tca gcc 335
 Cys Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp Tyr Ser Ala
 100 105 110

ccc act cgc ggc cct gcg gct aat tgc cgt cgt tcc gct ttg cgt ggt 383
 Pro Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu Arg Gly
 115 120 125

ctc cct cct gtc gac cgc act tac tgt yty gat gg 418
 Leu Pro Pro Val Asp Arg Thr Tyr Cys Xaa Asp
 130 135

<210> 252

<211> 138

<212> PRT

<213> Hepatitis E virus

<400> 252

Gly Xaa Thr Thr Xaa Ile Glu Gln Ala Ala Leu Ala Ala Ala Asn Ser
 1 5 10 15

Ala Leu Ala Asn Ala Val Val Val Arg Pro Phe Leu Ser Arg Val Gln
 20 25 30

Thr Glu Ile Leu Ile Asn Leu Met Xaa Pro Arg Xaa Leu Val Phe Arg
 35 40 45

Pro Glu Val Leu Trp Asn His Pro Ile Gln Arg Val Ile His Asn Glu
 50 55 60

Leu Glu Gln Tyr Cys Arg Thr Arg Ala Gly Arg Cys Leu Glu Val Gly
 65 70 75 80

Ala His Pro Arg Ser Ile Asn Asp Asn Pro Asn Val Xaa His Arg Cys
 85 90 95

Phe Leu Arg Pro Val Gly Arg Asp Val Gln Arg Trp Tyr Ser Ala Pro
 100 105 110

Thr Arg Gly Pro Ala Ala Asn Cys Arg Arg Ser Ala Leu Arg Gly Leu
 115 120 125

Pro Pro Val Asp Arg Thr Tyr Cys Xaa Asp
 130 135

<210> 253

<211> 197

<212> DNA

<213> Hepatitis E virus

<220>

<223> Ar2o2-f7702.pat

<220>

<221> CDS

<222> (2)..(196)

<400> 253

g aca gaa ttr att tcg tcg gct ggg ggt cag ttg ttt tac tcc cgc cct 49

Thr Glu Xaa Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro

1 5 10 15

gtc gtc tca gcc aat ggc gag ccg act gtt aag ttg tat aca tct gtg 97

Val Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val

20 25 30

gag aat gcg cag cag gat aaa gga atc gcc atc cca cac gac ata gat 145

Glu Asn Ala Gln Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp

35 40 45

ctg ggc gat tcc cgt gtg gtt att cag gat tat gay aac car cay gaa 193

Leu Gly Asp Ser Arg Val Val Ile Gln Asp Tyr Xaa Asn Xaa Xaa Glu

50 55 60

caa g

197

Gln

65

<210> 254

<211> 65

<212> PRT

<213> Hepatitis E virus

<400> 254

Thr Glu Xaa Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro

1 5 10 15

Val Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val

20 25 30

Glu Asn Ala Gln Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp

35 40 45

Leu Gly Asp Ser Arg Val Val Ile Gln Asp Tyr Xaa Asn Xaa Xaa Glu

50 55 60

Gln

65

<210> 255

<211> 23

<212> DNA

<213> Hepatitis E virus

<220>

<223> HEVConsORF 1N-a1

<400> 255

ccrtcrarrc artagggtgcg gtc

23

<210> 256
<211> 25
<212> DNA
<213> Hepatitis E virus

<220>
<223> HEVConsORF 2N-a1

<400> 256
cytgytcrtg ytggttrtca taatc

25

<210> 257
<211> 21
<212> DNA
<213> Hepatitis E virus

<220>
<223> HEVConsORF 1N-s2

<400> 257
cygccytkgc gaatgctgtg g

21

<210> 258
<211> 25
<212> DNA
<213> Hepatitis E virus

<220>
<223> HEVConsORF 2N-a2

<400> 258
gytcrtgytg rtrtrcataa tcctg

25